

```

XX      SQ      Sequence      32 AA;
XX      Query Match      59.3%; Score 64; DB 21; Length 32;
XX      Best Local Similarity 72.2%; Pred. No. 0.0058;
XX      Matches 13; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
OY      1 KMKLFKKIGIGKFLHSNK 18
XX      1 KMKLFKKIGIGKFLHSNK 18
Db      1 KMKLFKKIGIGKFLHSNK 18

RESULT 4
AAW97408
ID      AAW97408 standard; Protein; 39 AA.
XX
XX      AAW97408;
XX
XX      17-MAY-1999 (first entry)
XX
XX      Lytic peptide magainin 2+S.
XX
XX      magainin; transgenic; polinsetta; in vitro regeneration;
XX      resistance; ethylene; flower colour.
XX
XX      Synthetic.
XX
XX      WO9906566-A1.
XX
XX      11-FEB-1999.
XX
XX      31-JUL-1998; 98WO-US15917.
XX
XX      31-JUL-1997; 97US-0903944.
XX
XX      (SANF-) SANFORD SCI INC.
XX
XX      Blowers A, Chou T, Eisenreich R, Sanford J, Smith F;
XX      Van Eck J;
XX      WPI; 1999-153796/13.
XX      N-PSDB; AAX16060.
XX
XX      In vitro regeneration of Polinsetta - using embryo-induction medium
XX      containing casein hydrolase, used to produce transgenic plants
XX      with, e.g. increased disease resistance, controllable flower colour,
XX      altered habit and flowering time
XX
XX      Example 2; Page 51; 112pp; English.
XX
XX      The present sequence represents an lytic peptide which is used in the
XX      course of the invention to produce transgenic polinsetta plants with
XX      increased resistance to pathogens. The specification describes methods
XX      for the in vitro regeneration of Polinsetta. Transgenic polinsetta can
XX      now be produced with e.g. increased resistance to pathogens or insects;
XX      controllable flower colour; modified habit (more compact form, earlier
XX      flowering) and/or increased resistance to ethylene.
XX
XX      SQ      Sequence      39 AA;
XX
XX      Query Match      59.3%; Score 64; DB 20; Length 39;
XX      Best Local Similarity 92.3%; Pred. No. 0.0071;
XX      Matches 12; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
OY      8 IGIGKFLHSKKF 20
XX      :|||||
XX      16 MGIGKFLHSKKF 28
Db      16 MGIGKFLHSKKF 28

RESULT 5
AAW99125
ID      AAW99125 standard; peptide; 39 AA.
XX

```

```

AC      AAW99125;
XX
XX      17-MAY-1999 (first entry)
XX
XX      Extracellular localised antimicrobial magainin 2 peptide.
XX
XX      Antimicrobial peptide; resistance; plant pathogen; transgenic plant;
XX      powdery mildew.
XX
XX      Synthetic.
XX
XX      WO9906564-A1.
XX
XX      11-FEB-1999.
XX
XX      31-JUL-1998; 98WO-US16032.
XX
XX      31-JUL-1997; 97US-0054315.
XX
XX      (SANF-) SANFORD SCI INC.
XX
XX      Blowers AD, Sanford J, Smith F, Van Eck J;
XX      WPI; 1999-153794/13.
XX      N-PSDB; AAX19274.
XX
XX      New antimicrobial peptides - comprising magainin and PGL classes of
XX      peptides having an N-terminal methionine, used particularly for
XX      providing resistance in plants to pathogens
XX
XX      Claim 5; Page 40; 52pp; English.
XX
XX      The present sequence represents an antimicrobial peptide. The present
XX      invention also describes a method of conferring microbial resistance on
XX      a plant cell comprising transforming the cell to express two peptides
XX      selected from the magainin and PGL classes, where the peptides are
XX      separately compartmentalised. The antimicrobial peptide is effective
XX      against phytopathogenic microbes including bacteria, fungi, and
XX      phytoplasma. It can also be used to target other plant pests, such as
XX      nematodes and viruses. The nucleic acid sequence encoding the peptide
XX      can be used to transform plants to provide antimicrobial resistance. The
XX      method is especially useful in preventing powdery mildew. The method
XX      allows for levels of peptide expression to provide resistance to
XX      phytopathogenic microbes without otherwise affecting the plant cells.
XX
XX      SQ      Sequence      39 AA;
XX
XX      Query Match      59.3%; Score 64; DB 20; Length 39;
XX      Best Local Similarity 92.3%; Pred. No. 0.0071;
XX      Matches 12; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
OY      8 IGIGKFLHSKKF 20
XX      :|||||
XX      16 MGIGKFLHSKKF 28
Db      16 MGIGKFLHSKKF 28

RESULT 6
AAW67472
ID      AAW67472 standard; peptide; 19 AA.
XX
XX      AAW67472;
XX
XX      12-MAY-2000 (first entry)
XX
XX      Magainin 2 partial peptide fragment.
XX
XX      Antimicrobial; styelin; microbial degradation; bactericide; viricide;
XX      fungicide; plant protection; styelin A; magainin 2.
XX
XX      Unidentified.
XX
XX      CA2228730-A1.
XX

```

PD 24-JUN-1999.  
 XX  
 PF 14-APR-1998; 98CA-2228730.  
 XX  
 PR 24-DEC-1997; 97US-0068802.  
 PR 20-JAN-1998; 98US-0072885.  
 PR 18-FEB-1998; 98US-0075026.  
 XX  
 PA (REGC ) UNIV CALIFORNIA.  
 XX  
 PI Zhao C, Lee I, Lehrer RI;  
 DR WPI; 2000-161486/15.  
 XX  
 PT Antimicrobial styelin peptides isolated from *Styela clava* useful for  
 preserving materials vulnerable to microbial degradation and for  
 PT protecting plants and animals against pathogenic microbes -  
 XX  
 PS Example 2; Page 22; 35pp; English.  
 XX  
 CC The invention relates to antimicrobial peptides (styelins) (and  
 CC acetylated forms, salts, amides and esters of styelins), isolated from  
 CC *Styela clava* and characterized by specific patterns of basic and  
 CC hydrophobic amino acid side chains which display a broad spectrum of  
 CC antimicrobial activities. The peptides are of a specified formula and  
 CC display a wide range of antimicrobial activities and are therefore useful  
 CC for preserving materials susceptible to microbial degradation, for  
 CC protecting plants against bacterial infection and in the therapeutic and  
 CC prophylactic protection of animals against bacteria, fungi and viruses.  
 CC The compounds may also be used as standards in antimicrobial assays and  
 CC as affinity ligands for absorption of counterpart structures in microbes,  
 CC including viruses. The present sequence represents a partial fragment of  
 CC magainin 2, used in comparison studies.  
 CC  
 SQ Sequence 19 AA;  
 QY  
 Db 9 GIGKFLHSARKF 20  
 1 GIGKFLHSARKF 12  
 Db  
 RESULT 7  
 AAP71208  
 ID AAP71208 standard; Protein; 22 AA.  
 XX  
 AC AAP71208;  
 XX  
 DT 17-DEC-2001 (updated)  
 DT 01-MAY-1991 (first entry)  
 XX  
 DE Magainin III polypeptide.  
 XX  
 KM Magainin; antimicrobial; antibiotic.  
 XX  
 OS synthetic.  
 XX  
 PN USN7021493-N.  
 XX  
 PD 08-SEP-1987.  
 XX  
 PF 04-MAR-1987; 87US-0021493.  
 XX  
 PR 04-MAR-1987; 87US-0021493.  
 XX  
 PA (USSH ) US DEPT HEALTH & HUMAN.  
 XX  
 PI Zaslloff MA;  
 XX  
 DR WPI; 1987-334845/47.

XX  
 PT New antimicrobial polypeptide cpds. - are active against gram  
 PT positive and negative bacteria, fungi and protozoa  
 XX  
 PS Disclosure; page 2; 11pp; English.  
 XX  
 CC This magainin III polypeptide has a mol. wt. of 25kd or less, is water  
 CC sol. and non-cytolytic to animal cells, incl. RBCs, and is amphiphilic.  
 CC It is an active antimicrobial agent being effective against both  
 CC gram positive and -negative bacteria. It is also effective against  
 CC fungi and protozoa. It is therefore useful as an antibiotic.  
 CC See also AAP71206-07.  
 CC (Note: Revised entry submitted to correct the patent number format of  
 CC US Government-owned NPL applications to prevent clashes with ongoing US  
 CC granted patent numbers. For further information please visit the Derwent  
 CC web site at [www.derwent.com/dwpi/updates/ntis-us.html](http://www.derwent.com/dwpi/updates/ntis-us.html).)  
 CC  
 SQ Sequence 22 AA;  
 QY  
 Db 9 GIGKFLHSARKF 20  
 1 GIGKFLHSARKF 12  
 Db  
 RESULT 8  
 AAP82736  
 ID AAP82736 standard; protein; 22 AA.  
 XX  
 AC AAP82736;  
 XX  
 DT 17-DEC-2001 (updated)  
 DT 10-DEC-1990 (first entry)  
 XX  
 DE Magainin polypeptide (III) as bioactive cpds.  
 XX  
 KM Magainin polypeptide; cytotoxic agent; spermicidal agent;  
 KM antimicrobial agent; therapeutic; antitumour activity; contraceptive;  
 KM leukaemia; AIDS; HTLV III.  
 XX  
 OS synthetic.  
 XX  
 PN USN7076734-N.  
 XX  
 PD 23-FEB-1988.  
 XX  
 PF 23-JUL-1987; 87US-0076734.  
 XX  
 PR 23-JUL-1987; 87US-0076734.  
 PR 04-MAR-1987; 87US-0021493.  
 PR 05-AUG-1987; 87US-0081793.  
 XX  
 PA (USSH ) US DEPT HEALTH & HUMAN.  
 XX  
 PI Zaslloff MA;  
 XX  
 DR WPI; 1988-105363/15.  
 XX  
 PT New Magainin polypeptide(s) - active against Gram-positive and  
 PT -negative bacteria, fungi, viruses and protozoa and as cytotoxic  
 PT and spermicidal agents  
 XX  
 PS Disclosure; ; pp; English.  
 XX  
 CC To GI is attached NH2 and to N22 is attached OH.  
 CC Magainins are a class of substantially pure, homogeneous  
 CC polypeptides composed exclusively of about 25 amino acids, having a  
 CC mol. wt. of about 2500 or less, being water soluble at a concentration  
 CC of greater than 5 mg/ml at neutral pH or in an aq. solution of  
 CC physiologic ionic strength. They are amphiphilic surface-seeking and have

CC a broad spectrum of properties at physiologic ionic strength and pH.  
CC The polypeptide is a microbial agent capable of inhibiting the  
CC growth or proliferation of gram-positive and gram-negative bacteria,  
CC fungi, virus and protozoan species. They are also useful as  
CC therapeutic cytotoxic agents. This includes antitumor activity.  
CC They are useful as spermicides and as a contraceptive agent.  
CC Megalnin is active against HTLV III in vitro and has the potential  
CC to reduce the infectivity of the AIDS virus indicating its use as a  
CC sterilant against the virus or as a chemotherapeutic agent for treating  
CC AIDS.  
CC See also AAP82734-37.  
CC (Note: Revised entry submitted to correct the patent number format of  
CC US Government-owned NRTS applications to prevent clashes with ongoing US  
CC granted patent numbers. For further information please visit the Derwent  
CC web site at [www.derwent.com/dwpi/updates/nrtis-us.html](http://www.derwent.com/dwpi/updates/nrtis-us.html).)  
SQ Sequence 22 AA;

Query Match 58.3%; Score 63; DB 9; Length 22;  
Best Local Similarity 100.0%; Pred. No. 0.0057;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 9 GIGKFLHSAKKE 20  
| | | | | | | | | | | | | | | | | | | | | |  
Db 1 GIGKFLHSAKKE 12

## RESULT 9

AAR05732  
ID AAR05732 standard; protein; 22 AA.

XX AAR05732;

DT 05-NOV-1990 (first entry)

DE Deletion analogue of Megalnin II.

KW Megalnin II; antibiotic; spermicide; cancer; tumor;

KW fungicide; preservative.

OS Synthetic.

PN WO9006129-A.

PD 14-JUN-1990.

PF 05-DEC-1989; 89WO-0005499.

PR 09-DEC-1988; 88US-0281981.

PA (SCRT-) SCRIPPS CLINIC AND RE.

PI Cuervo JH, Houghten RA;

DR WPI; 1990-209623/27.

PT Deletion analogues of Megalnin I and II peptide(s) -  
PT having high activity against microorganisms with low haemolytic  
PT action against human red blood cells.

PS Claim 24; Page 24; 28pp; English.

CC Analogues have the same activity as Megalnin I and II coupled with  
CC insignificant haemolytic action against human red blood cells. Useful  
CC as antibiotics, fungicides, spermicides, in destroying tumors and  
CC as preservatives and in sterilisation of food etc.

SQ Sequence 22 AA;

Query Match 58.3%; Score 63; DB 11; Length 22;  
Best Local Similarity 100.0%; Pred. No. 0.0057;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 9 GIGKFLHSAKKE 20  
| | | | | | | | | | | | | | | | | | | | | |  
Db 1 GIGKFLHSAKKE 12

## RESULT 10

AAR05733  
ID AAR05733 standard; protein; 22 AA.

XX AAR05733;

DT 05-NOV-1990 (first entry)

DE Deletion analogue of Megalnin II.

KW Megalnin II; antibiotic; spermicide; cancer; tumor;

KW fungicide; preservative.

OS Synthetic.

PN WO9006129-A.

PD 14-JUN-1990.

PF 05-DEC-1989; 89WO-0005499.

PR 09-DEC-1988; 88US-0281981.

PA (SCRT-) SCRIPPS CLINIC AND RE.

PI Cuervo JH, Houghten RA;

DR WPI; 1990-209623/27.

PT Deletion analogues of Megalnin I and II peptide(s) -  
PT having high activity against microorganisms with low haemolytic  
PT action against human red blood cells.

PS Claim 25; Page 24; 28pp; English.

CC Analogues have the same activity as Megalnin I and II coupled with  
CC insignificant haemolytic action against human red blood cells. Useful  
CC as antibiotics, fungicides, spermicides, in destroying tumors and  
CC as preservatives and in sterilisation of food etc.

SQ Sequence 22 AA;

Query Match 58.3%; Score 63; DB 11; Length 22;  
Best Local Similarity 100.0%; Pred. No. 0.0057;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 9 GIGKFLHSAKKE 20  
| | | | | | | | | | | | | | | | | | | | | |  
Db 1 GIGKFLHSAKKE 12

## RESULT 11

AAR05734  
ID AAR05734 standard; protein; 22 AA.

XX AAR05734;

DT 05-NOV-1990 (first entry)

DE Deletion analogue of Megalnin II.

KW Megalnin II; antibiotic; spermicide; cancer; tumor;

KW fungicide; preservative.

OS Synthetic.

PN WO9006129-A.

```

PD 14-JUN-1990.
XX
PF 05-DEC-1989; 89WO-0005499.
XX
PR 09-DEC-1988; 88US-0281981.
XX
PA (SCRI-) SCRIPPS CLINIC AND RE.
XX
PI Cuervo JH, Houghten RA;
XX
DR WPI; 1990-209623/27.
XX
PT Deletion analogues of Magainin I and II peptide(s) -
PT having high activity against microorganisms with low haemolytic
XX action against human red blood cells.
XX
PS Claim 26; Page 24; 28pp; English.
XX
CC Analogues have the same activity as Megainin I and II coupled with
CC insignificant haemolytic action against human red blood cells. Useful
CC as antibiotics, fungicides, spermicides, in destroying tumors and
CC as preservatives and in sterilisation of food etc.
XX
SQ Sequence 22 AA;

Query Match 58.3%; Score 63; DB 11; Length 22;
Best Local Similarity 100.0%; Pred. No. 0.0057;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 9 GIGKFLHSAKKE 20
   |||||
Db 1 GIGKFLHSAKKE 12

RESULT 12
AAR05731
ID AAR05731 standard; protein; 22 AA.
XX
AC AAR05731;
XX
DT 05-NOV-1990 (first entry)
XX
DE Deletion analogue of Magainin II.
XX
KW Magainin II; antibiotic; spermicide; cancer; tumor;
KW fungicide; preservative.
XX
OS Synthetic.
XX
PN WO9006129-A.
XX
PD 14-JUN-1990.
XX
PF 05-DEC-1989; 89WO-0005499.
XX
PR 09-DEC-1988; 88US-0281981.
XX
PA (SCRI-) SCRIPPS CLINIC AND RE.
XX
PI Cuervo JH, Houghten RA;
XX
DR WPI; 1990-209623/27.
XX
PT Deletion analogues of Magainin I and II peptide(s) -
PT having high activity against microorganisms with low haemolytic
XX action against human red blood cells.
XX
PS Claim 23; Page 24; 28pp; English.
XX
CC Analogues have the same activity as Megainin I and II coupled with
CC insignificant haemolytic action against human red blood cells. Useful
CC as antibiotics, fungicides, spermicides, in destroying tumors and
CC as preservatives and in sterilisation of food etc.

```

```

XX
SQ Sequence 22 AA;

Query Match 58.3%; Score 63; DB 11; Length 22;
Best Local Similarity 100.0%; Pred. No. 0.0057;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 9 GIGKFLHSAKKE 20
   |||||
Db 1 GIGKFLHSAKKE 12

RESULT 13
AAR10658
ID AAR10658 standard; Protein; 22 AA.
XX
AC AAR10658;
XX
DT 18-APR-1991 (first entry)
XX
DE Magainin II analogue (XII).
XX
KW Magainin I; magainin II; substitution analogue; spermicide;
KW antibiotic; antiviral; antitumour.
XX
OS Synthetic.
XX
FT Key Location/Qualifiers
FT Modified-site 18..18
   /label= E-F-moc-lysine

WO9100869-A.
24-JAN-1991.
28-JUN-1990; 90WO-U003675.
07-JUL-1989; 89US-0376754.
(SCRI-) SCRIPPS CLINIC & RE.
Houghten RA, Cuervo JH;
WPI; 1991-051314/07.

Substitution analogues of magainin I and II - having increased
biological activity, useful as spermicide, antibiotic, antiviral
and antitumour agents

Claim 26; Page 55; 58pp; English.

The substitution analogue has biological activity equal to or
greater than the parent Magainin II peptide. The analogue is
effective as antibiotic and can be used to inhibit, prevent or
destroy the growth or proliferation of microbes such as Gram-
CC positive and Gram-negative bacteria, fungi, viruses and protozoa.
CC It can also be used as a spermicide to inhibit, prevent or
CC destroy the motility of sperm and as anti-tumour agent to inhibit
CC the growth of or destroy tumours.
CC See also AAR10639-47, AAR10657-63 and AAR10999.
XX
SQ Sequence 22 AA;

Query Match 58.3%; Score 63; DB 12; Length 22;
Best Local Similarity 100.0%; Pred. No. 0.0057;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 9 GIGKFLHSAKKE 20
   |||||
Db 1 GIGKFLHSAKKE 12

RESULT 14

```

```

AARI0641
ID AARI0641 standard; Protein; 22 AA.
XX
XX AARI0641;
AC
XX
XX 18-APR-1991 (first entry)
DT
XX
XX Magalain II analogue (IV).
DE
XX
XX Magalain I; magalain II; substitution analogue; spermicide;
KW antibiotic; antiviral; antitumour.
XX
XX Synthetic.
OS
XX
XX Key Location/Qualifiers
FH MISC-difference 18..18
FT /label= G, A
XX
XX WO9100869-A.
XX
XX 24-JAN-1991.
PD
XX
XX 28-JUN-1990; 90MO-U003675.
PF
XX
XX 07-JUL-1989; 89US-0376754.
PR
XX
XX (SCRI-) SCRIPPS CLINIC & RE.
PA
XX
XX Houghten RA, Cuervo JH;
PI
XX
XX WPI; 1991-051314/07.
DR
XX
XX Substitution analogues of magalain I and II - having increased
PT biological activity, useful as spermicide, antibiotic, antiviral
PT and antitumour agents
XX
XX
XX Claim 26; Page 55; 58pp; English.
XX
XX
XX The substitution analogue has biological activity equal to or
CC greater than the parent Magalain II peptide. The analogue is
CC effective as antibiotic and can be used to inhibit, prevent or
CC destroy the growth or proliferation of microbes such as Gram-
CC positive and Gram-negative bacteria, fungi, viruses and protozoa.
CC It can also be used as a spermicide to inhibit, prevent or
CC destroy the motility of sperm and as anti-tumour agent to inhibit
CC the growth of or destroy tumours.
CC See also AARI0639-47, AARI0657-63 and AARI0999.
CC
XX
XX
SQ Sequence 22 AA;

Query Match 58.3%; Score 63; DB 12; Length 22;
Best Local Similarity 100.0%; Pred. No. 0.0057;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 GIGKFLHSARKF 20
   |||||
   1 GIGKFLHSARKF 12

Db

RESULT 15
AARI0642
ID AARI0642 standard; Protein; 22 AA.
XX
XX AARI0642;
AC
XX
XX 18-APR-1991 (first entry)
DT
XX
XX Magalain II analogue (V).
DE
XX
XX Magalain I; magalain II; substitution analogue; spermicide;
KW antibiotic; antiviral; antitumour.
XX
XX Synthetic.
OS

```

```

XX
XX Key Location/Qualifiers
FH Modified-site 22..22
FT /note= "D-Serine"
XX
XX WO9100869-A.
XX
XX 24-JAN-1991.
PD
XX
XX 28-JUN-1990; 90MO-U003675.
PF
XX
XX 07-JUL-1989; 89US-0376754.
PR
XX
XX (SCRI-) SCRIPPS CLINIC & RE.
PA
XX
XX Houghten RA, Cuervo JH;
PI
XX
XX WPI; 1991-051314/07.
DR
XX
XX Substitution analogues of magalain I and II - having increased
PT biological activity, useful as spermicide, antibiotic, antiviral
PT and antitumour agents
XX
XX
XX Claim 26; Page 55; 58pp; English.
XX
XX
XX The substitution analogue has biological activity equal to or
CC greater than the parent Magalain II peptide. The analogue is
CC effective as antibiotic and can be used to inhibit, prevent or
CC destroy the growth or proliferation of microbes such as Gram-
CC positive and Gram-negative bacteria, fungi, viruses and protozoa.
CC It can also be used as a spermicide to inhibit, prevent or
CC destroy the motility of sperm and as anti-tumour agent to inhibit
CC the growth of or destroy tumours.
CC See also AARI0639-47, AARI0657-63 and AARI0999.
CC
XX
XX
SQ Sequence 22 AA;

Query Match 58.3%; Score 63; DB 12; Length 22;
Best Local Similarity 100.0%; Pred. No. 0.0057;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 GIGKFLHSARKF 20
   |||||
   1 GIGKFLHSARKF 12

Db

Search completed: June 6, 2003, 13:23:38
Job time : 31.5 secs

```



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 6, 2003, 13:22:11 ; Search time 12.5 Seconds  
(without alignments)  
153.815 Million cell updates/sec.

Title: US-10-081-418-1

Perfect score: 108

Sequence: 1 KMKLFKRGIGKFLHSAKFF 20

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: p1r1:\*  
2: p1r2:\*  
3: p1r3:\*  
4: p1r4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	63	58.3	303	2 A28620	magalainin precursor
2	48	44.4	376	2 I40728	recombination prot
3	47.5	44.0	237	2 C75035	arginase related p
4	47.5	44.0	427	2 A84155	hypothetical prote
5	47.5	44.0	491	2 T50950	hypothetical prote
6	47	43.5	1562	2 T43022	ATP-binding multid
7	46	42.6	412	2 S22474	phycoene synthase
8	46	42.6	412	2 A42102	phycoene synthase
9	46	42.6	666	2 G90566	hypothetical prote
10	45	41.7	64	1 CKMKA	cecropin A precursor
11	45	41.7	925	2 AE3628	nitrate reductase
12	44	40.7	205	2 C89957	hypothetical prote
13	44	40.7	353	2 S65297	probable membrane
14	44	40.7	512	2 H70662	probable p1cA prot
15	43	39.8	35	1 CKAOBP	cecropin B - chine
16	43	39.8	170	2 A75513	DNA topology modu
17	43	39.8	287	2 C71135	probable dihydrol
18	43	39.8	305	2 T27930	hypothetical prote
19	43	39.8	341	2 T18873	hypothetical prote
20	43	39.8	565	1 VHWIC8	nucleoprotein - in
21	43	39.8	589	2 T24378	hypothetical prote
22	43	39.8	620	2 F64408	coenzyme F420 hydr
23	43	39.8	662	2 UC7803	penicillin-binding
24	43	39.8	787	2 C75068	probable beta-gala
25	43	39.8	820	2 T26272	hypothetical prote
26	42.5	39.4	236	2 H82892	conserved hypothet
27	42	38.9	62	1 CKMKB	cecropin B precursor
28	42	38.9	62	1 CKMRA	cecropin A precursor
29	42	38.9	187	2 E84448	hypothetical prote

30	42	38.9	190	2 A64039	hypothetical prote
31	42	38.9	204	2 D69097	hypothetical prote
32	42	38.9	205	2 C95351	Vi-B5 type IV secr
33	42	38.9	325	2 A56406	casein kinase I (E
34	42	38.9	337	2 A57011	protein kinase CK1
35	42	38.9	337	2 S46254	hypothetical prote
36	42	38.9	380	2 T19096	probable phenylala
37	42	38.9	544	2 S75388	threonine ammonia-
38	42	38.9	600	2 T39516	hypothetical prote
39	42	38.9	5105	2 T32650	ubiquitinone/menaqu
40	41.5	38.4	248	2 F72262	A/G-specific adenil
41	41.5	38.4	339	2 B81258	serine/threonine/t
42	41.5	38.4	410	1 S28615	argininosuccinate
43	41.5	38.4	459	2 F84935	AAA family ATPase
44	41.5	38.4	809	2 T40537	hypothetical prote
45	41	38.0	106	2 T05199	

#### ALIGNMENTS

##### RESULT 1

A28620  
Magalainin precursor - African clawed frog

N:Contains: magalainin 1; magalainin 2

C:Species: Xenopus laevis (African clawed frog)

C>Date: 31-Dec-1988 #sequence\_revision 31-Dec-1988 #text\_change 18-Aug-2000

C:Accession: A28620; A29771

R:Berry, A.S.; Foulter, L.; Williams, D.H.; Nutkins, J.C.; Giovannini, M.G.; Moore, C

J. Biol. Chem. 263, 5745-5751, 1988

A:Title: The cDNA sequence coding for prepro-PGS (prepro-magalainin) and aspects of th

A:Reference number: A28620; PMID:88186892; PMID:2833514

A:Accession: A28620

A:Molecule type: mRNA

A:Residues: 1-303 <TER>

A:References: GB:J03193; NID:9214654; PIDN:AAA49930.1; PID:9214655

R:Zasloff, M.

Proc. Natl. Acad. Sci. U.S.A. 84, 5449-5453, 1987

A:Title: Magalainin, a class of antimicrobial peptides from Xenopus skin: Isolation, c

A:Reference number: A29771; PMID:87261003; PMID:3299384

A:Accession: A29771

A:Molecule type: mRNA

A:Residues: 6-73, 'Q', 'Y', 'S-158, 297-303 <ZAS>

C:Superfamily: magalainin precursor

Query Match 58.3%; Score 63; DB 2; Length 303;

Best Local Similarity 100.0%; Pred. No. 0.012; Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 GIGKFLHSAKFF 20

DB 83 GIGKFLHSAKFF 94

##### RESULT 2

I40728  
recombination protein recA - Corynebacterium glutamicum

N:Alternate names: recombinase A

C:Species: Corynebacterium glutamicum

C>Date: 16-Aug-1996 #sequence\_revision 16-Aug-1996 #text\_change 20-Apr-2001

C:Accession: I40728; S41939; I40727; S39258

R:Billman-Jacobe, H.

DNA Seq. 4, 403-404, 1994

A:Title: Nucleotide sequence of a recA gene from Corynebacterium glutamicum.

A:Reference number: I40728; PMID:95143581; PMID:7841463

A:Accession: I40728

A:Status: translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-376 <BILL>

A:Cross-references: EMBL:X77384; NID:9475063; PIDN:CAA54563.1; PID:9475064

A:Experimental source: strain AS019

R:Billman-Jacobe, H.

submitted to the EMBL Data Library, January 1994

A:Description: Nucleotide sequence of recA from *Corynebacterium glutamicum*.  
 A:Reference number: S41939  
 A:Accession: S41939  
 A:Molecule type: DNA  
 A:Residues: 1-102, 'DSF', 106-376 <BIL2>  
 A:Cross-references: EMBL:X77384  
 A:Experimental source: strain AS019  
 A:Ritzpatrick, R.; O'Donohue, M.; Joy, J.; Heery, D.M.; Duncan, L.K.  
 A:Title: Microbiol. Biotechnol. 42, 575-580, 1994  
 A:Title: Construction and characterisation of recA strains of *Corynebacterium glutamicum*  
 A:Reference number: I40727; MUID:55134369; PMID:7765733  
 A:Accession: I40727  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 118-200 <FIT2>  
 A:Cross-references: EMBL:X75085; NID:9474416; PIDN:CAA52977.1; PID:9474417  
 A:Ritzpatrick, R.; O'Donohue, M.; Joy, J.; Heery, D.M.; Duncan, L.K.  
 A:Submitted to the EMBL Data Library, September 1993  
 A:Description: Construction and characterisation of recA strains of *Corynebacterium glutamicum*  
 A:Reference number: S39258  
 A:Accession: S39258  
 A:Molecule type: DNA  
 A:Residues: 118-132, 'I', 134-169, 'GN', 172-190, 'S', 192-200 <FIT1>  
 A:Cross-references: EMBL:X75085  
 A:Experimental source: strain AS019  
 A:Genetics:  
 A:Gene: recA  
 C:Function:  
 A:Description: plays an essential role in homologous recombination, in induction of the C:Superfamily: recombination protein recA  
 C:Keywords: ATP; DNA binding; DNA recombination; DNA repair; nucleotide binding; P-loop; F/78-85/Region: nucleotide-binding motif A (P-loop)  
 F/152-157/Region: nucleotide-binding motif B  
 F/84/Binding site: ATP (Lys) #status predicted

Query Match 44.4%; Score 48; DB 2; Length 376;  
 Best Local Similarity 63.6%; Pred. No. 4.7;  
 Matches 7; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

OY 3 KLFKIGIGK 13  
 DB 334 KIRFKIGVKY 344

RESULT 3  
 C75035  
 arginase related protein PAB1539 [imported] - *Pyrococcus abyssi* (strain Orsay)  
 C:Species: *Pyrococcus abyssi*  
 C:Date: 20-Aug-1999 #sequence\_revision 20-Aug-1999 #text\_change 20-Jun-2000  
 C:Accession: C75035  
 R:anonymous, Genoscope  
 A:Submitted to the EMBL Data Library, July 1999  
 A:Description: *Pyrococcus abyssi* genome sequence: insights into archaeal chromosome structure  
 A:Reference number: A75001  
 A:Accession: C75035  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-237 <RAW>  
 A:Cross-references: GB:AJ248287; GB:AL06836; NID:95458657; PIDN:CA50176.1; PID:9545868  
 A:Experimental source: strain Orsay  
 C:Genetics:  
 A:Gene: PAB1539

Query Match 44.0%; Score 47.5; DB 2; Length 237;  
 Best Local Similarity 39.1%; Pred. No. 3.5;  
 Matches 9; Conservative 6; Mismatches 5; Indels 3; Gaps 1;

OY 1 KMKLF--KIGIGKFLHSARK 20  
 DB 135 KMKAYINLKOHGKGVIREARK 157

RESULT 4

A84155  
 hypothetical protein BH4041 [imported] - *Bacillus halodurans* (strain C-125)  
 C:Species: *Bacillus halodurans*  
 C:Date: 01-Dec-2000 #sequence\_revision 01-Dec-2000 #text\_change 15-Jun-2001  
 C:Accession: A84155  
 R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; H  
 Nucleic Acids Res. 28, 4317-4331, 2000  
 A:Title: Complete genome sequence of the alkaliphilic bacterium *Bacillus halodurans* a  
 A:Reference number: A83650; MUID:20512582; PMID:11058132  
 A:Accession: A84155  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-427 <STO>  
 A:Cross-references: GB:AP001520; GB:BA000004; NID:910176401; PIDN:BA07760.1; GSPDB:G  
 A:Experimental source: strain C-125  
 C:Genetics:  
 A:Gene: BH4041

Query Match 44.0%; Score 47.5; DB 2; Length 427;  
 Best Local Similarity 45.8%; Pred. No. 6.4;  
 Matches 11; Conservative 3; Mismatches 5; Indels 5; Gaps 1;

OY 1 KMKLFKIGIG----KFLHSARK 19  
 DB 217 KMKPFVKGVGIRKATLVETARK 240

RESULT 5  
 T50950  
 hypothetical protein B24P7.50 [imported] - *Neurospora crassa*  
 C:Species: *Neurospora crassa*  
 C:Date: 21-Jul-2000 #sequence\_revision 21-Jul-2000 #text\_change 08-Sep-2000  
 C:Accession: T50950  
 R:Schulte, U.; Aign, V.; Hohnsels, J.; Brandt, P.; Fattmann, B.; Holland, R.; Nyakatu  
 Submitted to the Protein Sequence Database, July 2000  
 A:Reference number: Z25286  
 A:Accession: T50950  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-491 <SCH>  
 A:Cross-references: EMBL:AL389890; GSPDB:GND0116; NCSP:B24P7.50  
 A:Experimental source: BAC clone B24P7; strain OR74A  
 C:Genetics:  
 A:Gene: NCSP:B24P7.50  
 A:Map position: 6  
 A:Insertions: 30/3; 235/2  
 C:Superfamily: *Neurospora crassa* hypothetical protein B24P7.50

Query Match 44.0%; Score 47.5; DB 2; Length 491;  
 Best Local Similarity 50.0%; Pred. No. 7.4;  
 Matches 10; Conservative 4; Mismatches 5; Indels 1; Gaps 1;

OY 2 WKLFK-KIGIGKFLHSARK 20  
 DB 269 WTLMKPCSGCHKYMSAEKF 288

RESULT 6  
 T43022  
 ATP-binding multidrug cassette transport protein - *Botryotinia fuckelliana*  
 C:Species: *Botryotinia fuckelliana*  
 C:Date: 11-Jan-2000 #sequence\_revision 11-Jan-2000 #text\_change 04-Mar-2000  
 C:Accession: T43022  
 R:Del Sorbo, G.; Andrade, A.C.; Van Nistelrooy, J. G.M.; Van Kan, J. A.L.; Balzi, E.; D  
 Mol. Gen. Genet. 254, 417-426, 1997  
 A:Title: Multidrug resistance in *Aspergillus nidulans* involves novel ATP-binding cass  
 A:Reference number: Z07910; MUID:9180695; PMID:9180695  
 A:Accession: T43022  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-1562 <DEL>  
 A:Cross-references: EMBL:Z68906; NID:e989397; PID:e219958; PIDN:CAA93142.1  
 A:Experimental source: strain SAS56; clone bc-atra



## C:Genetics:

A:introns: 188/3; 338/3; 655/3; 711/2; 814/1; 1356/3  
C:Superfamily: unassigned ATP-binding cassette proteins; ATP-binding cassette homology

## Query Match

Best Local Similarity 43.5%; Score 47; DB 2; Length 1562;  
Matches 8; Conservative 7; Mismatches 3; Indels 8; Gaps 1;

OY 1 KWK-----LFRKIGIGKFLHSAK 18  
DB 1497 KWRNFGIFGLIYQOLGVGLHGOR 1522

## RESULT 7

phytoene synthase (EC 2.5.1.-) - tomato

N:Alternate names: ripening protein 5

C:Species: Lycopersicon esculentum (tomato)

C:Date: 19-Mar-1997 #sequence,revision 18-Jul-1997 #text\_change 02-Mar-2001

C:Accession: S22474; S06321; S35155; S35154; S21981; S24967; S24968

R:Ray, J.; Moureau, P.; Bird, C.; Bird, A.; Grierson, D.; Maunders, M.; Truesdale, M.; E

Plant Mol. Biol. 19, 401-404, 1992

A:Title: Cloning and characterization of a gene involved in phytoene synthesis from toma

A:Reference number: S22474; MUID:92322971; PMID:1623189

A:Accession: S22474

A:Molecule type: DNA

A:Residues: 1-412 <RAY>

A:Cross-references: EMBL:X60441; GB:S39572; NID:919248; PIDN:CAA42969.1; PID:919249

A:Note: The authors did not translate the codon for residue 155

R:Ray, J.; Bird, C.; Maunders, M.; Grierson, D.; Schuch, W.

Nucleic Acids Res. 15, 10587, 1987

A:Title: Sequence of PTOM5, a ripening related cDNA from tomato.

A:Reference number: S06321; MUID:88096591; PMID:3697097

A:Accession: S06321

A>Status: not compared with conceptual translation

A:Molecule type: mRNA

A:Residues: 1-406, 'N', 408-412 <RAY1>

A:Cross-references: GB:X00521; NID:919340; PIDN:CAA68575.1; PID:919341

R:Frery, R.G.; Grierson, D.

Plant Mol. Biol. 22, 589-602, 1993

A:Title: Identification and genetic analysis of normal and mutant phytoene synthase gene

A:Reference number: S35154; MUID:93344508; PMID:8343597

A:Accession: S35155

A>Status: translation not shown

A:Molecule type: mRNA

A:Residues: 1-387, 'NMKDFSNFKSGKSGSNATTIVGLAPCE', 'TYIADRGPIGTF' <FRA1>

A:Cross-references: EMBL:X67144; NID:919346; PIDN:CAA47625.1; PID:919347

A:Experimental source: mutant rx

A:Accession: S35154

A>Status: translation not shown

A:Molecule type: mRNA

A:Residues: 1-107, 'VECLESDPLQTERGVSLPQVGVSGRGAT', 'RMPGGPGRRGAGCIGCPLIFC' <FRA2>

A:Cross-references: EMBL:X67143; NID:919396; PIDN:CAA47624.1; PID:919399

A:Experimental source: mutant r

C:Genetics:

A:Gene: Psy1

A:introns: 138/1; 155/1; 212/3; 291/2; 355/3

C:Superfamily: Mycobacterium marinum phytoene synthase

C:Keywords: chloroplast; transferase

OY 1 KWKLFKIGIGKFLHSAK 20

DB 322 KWRIFMK-----KQIHRAKRF 337

## Query Match

Best Local Similarity 42.6%; Score 46; DB 2; Length 412;  
Matches 9; Conservative 4; Mismatches 3; Indels 4; Gaps 1;

RESULT 8  
AA2102  
phytoene synthase (EC 2.5.1.-) peripheral plastid membrane - tomato  
C:Species: Lycopersicon esculentum (tomato)

C:Date: 04-Mar-1993 #sequence,revision 18-Nov-1994 #text\_change 02-Mar-2001

C:Accession: AA2102

R:Barley, G.E.; Viltanen, P.V.; Bacot, K.O.; Scolnik, P.A.

J. Biol. Chem. 267, 5036-5039, 1992

A:Title: A tomato gene expressed during fruit ripening encodes an enzyme of the carot

A:Reference number: AA2102; MUID:92184738; PMID:1544888

A:Accession: AA2102

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-412 <BAR>

A:Cross-references: GB:M84744; NID:9170415; PIDN:AAA34153.1; PID:9170416

A:Note: sequence extracted from NCBI backbone (NCBIN:87066, NCBIIP:87067)

C:Superfamily: Mycobacterium marinum phytoene synthase

C:Keywords: chloroplast; transferase

## Query Match

Best Local Similarity 42.6%; Score 46; DB 2; Length 412;  
Matches 9; Conservative 4; Mismatches 3; Indels 4; Gaps 1;

OY 1 KWKLFKIGIGKFLHSAK 20  
DB 322 KWRIFMK-----KQIHRAKRF 337

## RESULT 9

hypothetical protein MYPV\_4390 [imported] - Mycoplasma pulmonis (strain UAB CTRP)

C:Species: Mycoplasma pulmonis

C:Date: 24-May-2001 #sequence,revision 24-May-2001 #text\_change 03-Aug-2001

C:Accession: G90566

R:Chamhaud, I.; Hellig, R.; Ferris, S.; Barbe, V.; Samson, D.; Galisson, F.; Moszer,

Nucleic Acids Res. 29, 2145-2153, 2001

A:Title: The complete genome sequence of the murine respiratory pathogen Mycoplasma p

A:Reference number: A98512; MUID:21267165; PMID:11353084

A:Accession: G90566

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-666 <KUR>

A:Cross-references: GB:AL445566; PID:914089853; PIDN:CAAC13612.1; GSPDB:GN00153

A:Experimental source: strain UAB CTRP

C:Genetics:

A:Gene: MYPV\_4390

A:Genetic code: SGC3

## Query Match

Best Local Similarity 42.6%; Score 46; DB 2; Length 666;  
Matches 9; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

OY 1 KWKLFKIGIGKFLHSAK 19  
DB 350 KKHIFKIGLGFYSIFEK 368

## RESULT 10

CEKWA

N:Alternate names: cecropia moth

C:Species: Hyalophora cecropia (cecropia moth)

C:Date: 24-Sep-1981 #sequence,revision 04-Nov-1994 #text\_change 18-Jun-1999

C:Accession: A40420; S00208; A91121; A93260; A01769

R:Gundundson, G.H.; Lidholm, D.A.; Asling, B.; Gan, R.; Boman, H.G.

J. Biol. Chem. 266, 11510-11517, 1991

A:Title: The cecropin locus. Cloning and expression of a gene cluster encoding three

A:Reference number: A40420; MUID:91268009; PMID:1711035

A:Accession: A40420

A:Molecule type: DNA

A:Residues: 1-64 <GUD>

A:Cross-references: GB:M63845; NID:9159196; PIDN:AAA29185.1; PID:9159197

R:Lidholm, D.A.; Gundundson, G.H.; Xanthopoulos, K.G.; Boman, H.G.

FEBS Lett. 226, 8-12, 1987

A:Title: Insect immunity: cDNA clones coding for the precursor forms of cecropins A a

A:Reference number: S00208

A:Accession: S00208

A:Molecule type: mRNA  
A:Residues: 1-64 <LID>  
A:Cross-references: EMBL:X06672; NID:99413; PIDN:CAA29871.1; PID:9671870  
R:Hallmark, D.; Engstrom, A.; Bennich, H.; Kapur, R.; Boman, H.G.  
Eur. J. Biochem. 127, 207-217, 1982  
A:Title: Insect immunity: isolation and structure of cecropin D and four minor antibacterial peptides  
A:Reference number: A91121; MUID:83053366; PMID:7140755  
A:Accession: A91121  
A:Molecule type: protein  
A:Residues: 27-63 <HUL>  
A:Note: cecropin C lacks the carboxyl-terminal blocking group  
R:Steiner, H.; Hallmark, D.; Engstrom, A.; Bennich, H.; Boman, H.G.  
Nature 292, 246-248, 1981  
A:Title: Sequence and specificity of two antibacterial proteins involved in insect immunity  
A:Reference number: A93260; MUID:81245158; PMID:7019715  
A:Accession: A93260  
A:Molecule type: protein  
A:Residues: 27-63 <STE>  
R:Andreu, D.; Merrifield, R.B.; Steiner, H.; Boman, H.G.  
Proc. Natl. Acad. Sci. U.S.A. 80, 6475-6479, 1983  
A:Title: Solid-phase synthesis of cecropin A and related peptides.  
A:Reference number: A54960; MUID:84042494; PMID:6579533  
A:Contents: annotation; chemical synthesis  
A:Note: protein carboxyl-terminal composition and amidation determined by mass spectrometry  
C:Comment: Cecropins are a family of basic peptides having lytic and antibacterial activity  
C:Genetics:  
A:Introns: 34/3  
C:Superfamily: cecropin  
C:Keywords: amidated carboxyl end; antibacterial; hemolymph  
F:1-22/Domain: signal sequence #status predicted <SIG>  
F:23-26/Domain: propeptide #status predicted <PRO>  
F:27-63/Product: cecropin A #status experimental <MAT>  
F:63/Modified site: amidated carboxyl end (Lys) (amide in mature form from following gly)

Query Match 41.7%; Score 45; DB 1; Length 64;  
Best Local Similarity 100.0%; Pred. No. 2.4;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 KMKLFKKIKGKFLHSKKE 8  
Db 27 KMKLFKKIK 34

RESULT 11  
AE3628  
nitrate reductase (EC 1.7.99.4) [imported] - Brucella melitensis (strain 16M)  
C:Species: Brucella melitensis  
C:Date: 01-Feb-2002 #sequence\_revision 01-Feb-2002 #text\_change 03-Jun-2002  
C:Accession: AE3628  
R:DeLVecchio, V.G.; Kapral, V.; Redkar, R.J.; Patra, G.; Mujer, C.; Los, T.; Ivanova, R.; Mazur, M.; Goldsman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Letessier, Natl. Acad. Sci. U.S.A. 99, 443-448, 2002  
A:Title: The genome sequence of the facultative intracellular pathogen Brucella melitensis  
A:Reference number: AD3522; PMID:11756688  
A:Accession: AE3628  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-925 <KUR>  
A:Cross-references: GB:AE008918; PIDN:AAL54192.1; PID:g17985160; GSPDB:GN00191  
A:Experimental source: strain 16M  
C:Genetics:  
A:Gene: BME110950  
A:Map position: 11  
C:Keywords: oxidoreductase

Query Match 41.7%; Score 45; DB 2; Length 925;  
Best Local Similarity 50.0%; Pred. No. 37;  
Matches 10; Conservative 2; Mismatches 8; Indels 0; Gaps 0;

Oy 1 KMKLFKKIKGKFLHSKKE 20  
Db 72 KMKLFKKIKGKFLHSKKE 91

RESULT 12  
C89957  
hypothetical protein SA1548 [imported] - Staphylococcus aureus (strain N315)  
C:Species: Staphylococcus aureus  
C:Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 22-Oct-2001  
C:Accession: C89957  
R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oma, A.; Mizutani, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kato, C.; Sekimizu, K.  
Lancet 357, 1225-1240, 2001  
A:Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.  
A:Reference number: A89758; MUID:21311952; PMID:11418146  
A:Accession: C89957  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-205 <KUR>  
A:Cross-references: GB:BA000018; PID:g13701522; PIDN:BA842816.1; GSPDB:GN00149  
A:Experimental source: strain N315  
C:Genetics:  
A:Gene: SA1548

Query Match 40.7%; Score 44; DB 2; Length 205;  
Best Local Similarity 55.0%; Pred. No. 12;  
Matches 11; Conservative 1; Mismatches 8; Indels 0; Gaps 0;

Oy 1 KMKLFKKIKGKFLHSKKE 20  
Db 63 KMKLFKKIKGKFLHSKKE 82

RESULT 13  
S65297  
probable membrane protein YPL264c - Yeast (Saccharomyces cerevisiae)  
C:Species: Saccharomyces cerevisiae  
C:Date: 10-Dec-1994 #sequence\_revision 31-May-1996 #text\_change 19-Apr-2002  
C:Accession: S65297; S65318  
R:Duesterhoeft, A.; Floeth, M.; Fritze, M.; Hilbert, H.; Moestl, D.  
submitted to the Protein Sequence Database, May 1996  
A:Reference number: S65292  
A:Accession: S65297  
A:Molecule type: DNA  
A:Residues: 1-353 <DOE>  
A:Cross-references: EMBL:Z73620; NID:g1370544; PID:e246970; PID:g1370545; MIPS:YPL264  
A:Experimental source: strain S288C (AB972)  
R:Delius, H.; Hebling, U.  
submitted to the Protein Sequence Database, May 1996  
A:Reference number: S64967  
A:Accession: S65318  
A:Molecule type: DNA  
A:Residues: 1-353 <DEL>  
A:Cross-references: EMBL:Z73620; NID:g1370544; PID:e246970; PID:g1370545; MIPS:YPL264  
A:Experimental source: strain S288C (AB972)  
C:Genetics:  
A:Cross-references: SGD:S0006185  
A:Map position: 16L  
C:Keywords: transmembrane protein  
F:16-33/Domain: transmembrane #status predicted <TM1>  
F:54-70/Domain: transmembrane #status predicted <TM2>  
F:96-112/Domain: transmembrane #status predicted <TM3>  
F:118-134/Domain: transmembrane #status predicted <TM4>  
F:143-159/Domain: transmembrane #status predicted <TM5>  
F:189-205/Domain: transmembrane #status predicted <TM6>  
F:224-240/Domain: transmembrane #status predicted <TM7>

Query Match 40.7%; Score 44; DB 2; Length 353;  
Best Local Similarity 46.7%; Pred. No. 20;  
Matches 7; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Oy 1 KMKLFKKIKGKFLHSKKE 15  
Db 253 QWGLFNLIGISGFH 267

RESULT 14  
H70663

RESULT 14  
 H70662  
 probable plca protein - Mycobacterium tuberculosis (strain H37Rv)  
 C:Species: Mycobacterium tuberculosis  
 C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 20-Jun-2000  
 C:Accession: H70662  
 R:COLE, S.T.; BROSH, R.; PARKHILL, J.; GARNIER, T.; CHURCHER, C.; HARRIS, D.; GORDON, G.; COMOR, R.; DAVIES, R.; DEVLIN, K.; FELLWELL, T.; GENTLES, S.; HAMLIN, N.; HOLROYD, S.; REJLANDREEM, M.A.; ROGERS, J.; RUTTER, S.; SEEGER, K.; SKELTON, S.; SQUARES, S.  
 N:ature 333, 537-544, 1998  
 A:Authors: Soares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
 A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome  
 A:Reference number: A70500; MUID:98295987; PMID:9634230  
 A:Accession: H70662  
 A:Stictus: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-512 <COL>  
 A:Cross-references: GB:283860; GB:AL123456; NID:g3261681; PIDN:CAB06148.1; PID:g1781258  
 A:Experimental source: strain H37Rv  
 C:Genetics:  
 A:Gene: plca

Query Match	40.7%;	Score 44;	DB 2;	Length 512;
Best Local Similarity	33.3%;	Pred. NO. 30;		
Matches	5;	Conservative	8;	Mismatches 2;
				Indels 0;
				Gaps 0;

```
QY      2 WKLEFKKIGIGKFLHS 16
         ||::: |::|:::
Db      233 WKVYQNKGLGRFINT 247
```

**RESULT 15**

CKA0BP  
C:ecropin B - Chinese oak silkworm  
C:Species: *Antheraea pernyi* (Chinese oak silkworm)  
C:Date: 17-Dec-1982 #sequence\_revision 12-Apr-1996 #text\_change 07-May-1999  
C:Accession: A54775; A01771  
R:Craig, A.G.; Engstrom, A.; Benich, H.; Kamensky, I.  
Biomed. Environ. Mass Spectrom. 14, 669-673, 1987  
A:Title: Plasma desorption mass spectrometry coupled with conventional peptide sequencing  
A:Reference number: A54725; MUID:88108273; PMID:2562676  
A:Accession: A54725  
A:Molecule type: protein  
A:Residues: 1-35 <CR>  
R:Ou, X.; Steiner, H.; Engstrom, A.; Benich, H.; Boman, H.G.  
Eur. J. Biochem. 127, 219-224, 1982  
A:Title: Insect Immunology: Isolation and structure of ecropins B and D from pupae of the silkworm  
A:Reference number: A91122; MUID:83053368; PMID:6754375  
A:Accession: A01771  
A:Molecule type: protein  
A:Residues: 1-35, 'S' <QUX>  
A:Note: the carboxyl end is blocked and was not sequenced directly  
C:Superfamily: ecropin  
C:Keywords: amidated carboxyl end; antibacterial  
F:35/Modified site: amidated carboxyl end (neu) #status experimental

Query Match	39.8%;	Score 43;	DB 1;	Length 35;
Best Local Similarity	87.5%;	Pred. No. 2.7;		
Matches	7;	Conservative	0;	Indels 0;
				Gaps 0;

Qy	1	KWKLEKKI	8
		:	
Db	1	KWKIEKKI	8

Search completed: June 6, 2003, 13:25:26  
Job time : 14.5 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 6, 2003, 13:21:00 ; Search time 7 Seconds

(without alignments)  
118.504 Million cell updates/sec

Title: US-10-081-418-1

Perfect score: 108

Sequence: 1 KMKLFKKIGIGKFLHSKKF 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 08

Maximum Match 1008

Database: SwlssProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	63	58.3	303	1	MAGA_XENLA
2	46	44.4	376	1	RECA_CORGL
3	46	42.6	412	1	PSY1_LYCES
4	45	41.7	64	1	CECA_HYACE
5	44	40.7	520	1	PHLA_MCTU
6	43	39.8	35	1	CECB_ANTPE
7	43	39.8	341	1	YKLI_GAEEL
8	43	39.8	363	1	MTD3_STYHU
9	43	39.8	565	1	VNIC_INCCA
10	43	39.8	620	1	Y870_MERTJA
11	42.5	39.4	713	1	GALX_SCHPO
12	42	38.9	62	1	CECB_HYACE
13	42	38.9	63	1	CECA_BOMMO
14	42	38.9	190	1	YG31_HAEIN
15	42	38.9	325	1	KCIA_BOVIN
16	42	38.9	325	1	KCIA_RAT
17	42	38.9	337	1	KCIA_CHICK
18	42	38.9	337	1	KCIA_HUMAN
19	42	38.9	544	1	STPB_SULSO
20	42	38.9	644	1	NKX3_HUMAN
21	41.5	38.4	410	1	APRA_ARATH
22	41.5	38.4	459	1	ARLY_BUCAT
23	41	38.0	305	1	BLAI_HAEIN
24	41	38.0	336	1	KCIB_HAEIN
25	41	38.0	552	1	STI_SULAC
26	41	38.0	624	1	NKX3_RAT
27	41	38.0	645	1	NKX3_MOUSE
28	41	38.0	713	1	CNBA_HUMAN
29	41	38.0	791	1	POLG_DENIC
30	40.5	37.5	690	1	PCDB_DROME
31	40	37.0	35	1	CECA_BOMMO
32	40	37.0	63	1	CE3D_HYPCU
33	40	37.0	63	1	CECB_BOMMO

## ALIGNMENTS

34	40	37.0	208	1	MLEY_HUMAN	P14649	homo sapien
35	40	37.0	247	1	RECO_NEIGO	Q9zhy2	neisseria g
36	40	37.0	368	1	YI05_YEAST	P40577	saccharomyc
37	40	37.0	662	1	PM7_YEAST	O06644	saccharomyc
38	40	37.0	684	1	TC10_YEAST	P50273	saccharomyc
39	40	37.0	755	1	TR2M_AGR4	P04029	agrobacteri
40	40	37.0	755	1	TR2N_AGR4	P25017	agrobacteri
41	40	37.0	1096	1	PDPI_ARATH	Q91rz5	arabidopsis
42	40	37.0	1663	1	HAPF_HUMAN	O60229	homo sapien
43	40	37.0	1919	1	HAPF_RAT	P97924	rattus norv
44	40	37.0	3038	1	TRIO_HUMAN	O75962	homo sapien
45	40	37.0	3390	1	POLG_DEN3	P27915	d genome po

RESULT 1  
MAGA\_XENLA  
ID MAGA\_XENLA STANDARD: PRT; 303 AA.

AC P11006;

DT 01-JUL-1989 (Rel. 11, Created)

DT 01-JUL-1989 (Rel. 11, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Magalins precursor.

OS Xenopus laevis (African clawed frog).

OC Amphibia; Batrachia; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;

OX NCBI\_Taxid=8355;

RN [1]  
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.

RX MEDLINE=88186892; PubMed=2833514;

RA Terry A.S., Poulter L., Williams D.H., Nutkins J.C., Giovannini M.G.,

RT Moore C.H., Gibson B.W.;

RT "The cDNA sequence coding for prepro-PGS (prepro-magalins) and

RT aspects of the processing of this prepro-polypeptide.";

RL J. Biol. Chem. 263:5745-5751(1988).

[2]  
RP SEQUENCE OF 6-158 AND 297-303 FROM N.A., AND PARTIAL SEQUENCE.

RX MEDLINE=87261003; PubMed=3299384;

RA Zaslouff M.;

RT "Magalins, a class of antimicrobial peptides from Xenopus skin:

RT isolation, characterization of two active forms, and partial cDNA

RL Proc. Natl. Acad. Sci. U.S.A. 84:5449-5453(1987).

[3]  
RP SEQUENCE OF MAGALINS I AND II.

RX MEDLINE=92011794; PubMed=1717472;

RA Moore K.S., Bevilas C.L., Brasseur M.M., Tomassini N., Turner K.,

RT Eck H., Zaslouff M.;

RT "Antimicrobial peptides in the stomach of Xenopus laevis.";

RL J. Biol. Chem. 266:19851-19857(1991).

[4]  
RP STRUCTURE BY NMR OF MAGALIN II.

RX MEDLINE=94129391; PubMed=8298457;

RA Bechinger B., Zaslouff M., Opella S.J.;

RT "Structure and orientation of the antibiotic peptide magalin in

RT membranes by solid-state nuclear magnetic resonance spectroscopy.";

RL Protein Sci. 2:2077-2084(1993).

-1- FUNCTION: ANTIMICROBIAL PEPTIDES THAT INHIBIT THE GROWTH OF

CC OF PROTOZOA. MAGALINS ARE MEMBRANE LYtic AGENTS.

CC -1- TISSUE SPECIFICITY: SYNTHESIZED IN THE STOMACH AND STORED IN A

CC NOVEL GRANULAR MULTINUCLEATED CELL IN THE GASTRIC MUCOSA. IT IS

CC STORED AS ACTIVE, PROCESSED PEPTIDES IN LARGE GRANULES WITHIN

CC THE GRANULAR GLAND SECRETIONS OF THE SKIN.

CC -1- SIMILARITY: BELONGS TO THE MAGALIN FAMILY OF ANTIMICROBIAL

CC PEPTIDES.

CC -1- DATABASE: NMBE-protein Spotlight;

CC NOTE-Issue 7 of February 2001;

CC WWW="http://www.expasy.org/spotlight/articles/spl1007.html".

```

CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC
DR      EMBL: J03193; AAA4930.1; -
DR      PIR: A28620; A28620.
DR      PIR: A29771; A29771.
DR      PDB: 2MAG; 08-APR-98.
DR      InterPro: IPR001651; Gastrin.
DR      Pfam: PF00918; Gastrin; 1.
KM      Cleavage on pair of basic residues; Repeat; Amphibian skin;
KW      Antibiotic; Fungicide; Hemolysis; Signal; 3D-structure.
FT      SIGNAL 1 18
FT      PROPEP 19 26
FT      PEPTIDE 27 32
FT      PROPEP 33 36
FT      PEPTIDE 37 59
FT      PROPEP 62 72
FT      PEPTIDE 73 78
FT      PROPEP 79 82
FT      PEPTIDE 83 105
FT      PROPEP 108 118
FT      PEPTIDE 119 124
FT      PROPEP 125 128
FT      PEPTIDE 129 151
FT      PROPEP 154 164
FT      PEPTIDE 165 170
FT      PROPEP 171 174
FT      PEPTIDE 175 197
FT      PROPEP 200 210
FT      PEPTIDE 211 216
FT      PROPEP 217 220
FT      PEPTIDE 221 243
FT      PROPEP 246 256
FT      PEPTIDE 257 262
FT      PROPEP 263 266
FT      PEPTIDE 267 289
FT      PROPEP 292 303
FT      CONFLICT 74 74
SQ      SEQUENCE 303 AA; 33379 MW; E369BDBB033EAB0 CRC64;
      E -> Q (IN REF. 2).
      Query Match 58.3%; Score 63; DB 1; Length 303;
      Best Local Similarity 100.0%; Pred. No. 0.0068;
      Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY      9 GIGKFLHSAKKF 20
      |||||
DB      83 GIGKFLHSAKKF 94

```

```

RL      Submitted (SEP-1994) to the EMBL/GenBank/DBJ databases.
RN      [2]
RP      SEQUENCE FROM N.A.
RC      STRAIN-ATCC 13059 / AS019;
RX      MEDLINE=95143581; PubMed=7841463;
RA      Billman-Jacobe H.;
RT      "Nucleotide sequence of a recA gene from Corynebacterium glutamicum.";
RL      DNA Seq. 4:403-404(1994).
RN      [3]
RP      SEQUENCE FROM N.A.
RC      STRAIN-ATCC 13032 / DSM 20300 / NCIB 10025;
RA      Nakagawa S.;
RT      "Complete genomic sequence of Corynebacterium glutamicum ATCC 13032.";
RL      Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
RN      [4]
RP      SEQUENCE OF 118-200 FROM N.A.
RC      STRAIN-ATCC 13059 / AS019;
RX      MEDLINE=95143659; PubMed=7765733;
RA      Fitzpatrick R., O'Donohue M., Joy J., Heery D.M., Dunican L.K.;
RT      "Construction and characterization of recA mutant strains of
RT      Corynebacterium glutamicum and Brevibacterium lactofermentum.";
RL      Appl. Microbiol. Biotechnol. 42:575-580(1994).
CC      -1- FUNCTION: CAN CATALYZE THE HYDROLYSIS OF ATP IN THE PRESENCE OF
CC      SINGLE-STRANDED DNA, THE ATP-DEPENDENT UPTAKE OF SINGLE-STRANDED
CC      DNA BY DUPLEX DNA, AND THE ATP-DEPENDENT HYDRIDIZATION OF
CC      HOMOLOGOUS SINGLE-STRANDED DNAs. IT INTERACTS WITH LEXA CAUSING
CC      ITS ACTIVATION AND LEADING TO ITS AUTOCATALYTIC CLEAVAGE.
CC      -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC      -1- SIMILARITY: BELONGS TO THE RECA FAMILY.
CC
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC
CC      EMBL: U14965; AAD12743.1; -
CC      EMBL: X77384; CAA54563.1; -
CC      EMBL: AP005280; BAB99348.1; -
CC      EMBL: X75085; CAA52977.1; -
CC      HSP: P26345; IGI9.
CC      InterPro: IPR003593; AAA_ATPase.
CC      InterPro: IPR001553; RecA.
CC      Pfam: PF00154; recA; 1.
CC      PRINTS: PR00142; RECA.
CC      ProDom: PD000229; RecA; 1.
CC      SMART: SM00382; AAA; 1.
CC      PROSITE: PS00321; RECA_1; 1.
CC      PROSITE: PS50162; RECA_2; 1.
CC      PROSITE: PS50163; RECA_3; 1.
CC      DNA damage; DNA recombination; SOS response; ATP-binding; DNA-binding.
CC      NP BIND 78 85
CC      CONFLICT 2 2
CC      FT      A -> H (IN REF. 2).
CC      FT      SEQUENCE 376 AA; 40227 MW; 5535924F18549D5C CRC64;
      Query Match 44.4%; Score 48; DB 1; Length 376;
      Best Local Similarity 63.6%; Pred. No. 2.3;
      Matches 7; Conservative 4; Mismatches 0; Indels 0; Gaps 0;
OY      3 KLFKKIGIGKF 13
      |::|::|::|
DB      334 KLFKKIGVGY 344

```

```

RESULT 2
RECA_CORGL STANDARD: PRT; 376 AA.
AC      P42442;
DT      01-NOV-1995 (Rel. 32, Created)
DT      01-NOV-1995 (Rel. 32, Last sequence update)
DT      15-JUN-2002 (Rel. 41, Last annotation update)
DE      RecA protein (Recombinase A).
GN      RECA OR CGL1955.
OS      Corynebacterium glutamicum (Brevibacterium flavum).
OC      Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC      Corynebacterales; Corynebacteriaceae;
OC      Corynebacterium.
OX      NCBI_TaxID=1718;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN-ATCC 13059 / AS019;
RA      Kerins S.M., Fitzpatrick R., O'Donohue M., Dunican L.K.;

```

DE phytoene synthase 1, chloroplast precursor (EC 2.5.1.-) (Fruit  
 ripening specific protein PTOM5).  
 GN PSY1 OR PTOM5.  
 OS Lycopersicon esculentum (Tomato).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 OC Asteridae; euasterids I; Solanales; Solanaceae; Solanum.  
 ON NCBI\_TaxID=4081;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND CHARACTERIZATION.  
 RX MEDLINE=92184738; PubMed=1544888;  
 RA Bartley G.E., Viltanen P.V., Bacot K.O., Scolnik P.A.;  
 RT "A tomato gene expressed during fruit ripening encodes an enzyme of  
 the carotenoid biosynthesis pathway.";  
 RL J. Biol. Chem. 267:5036-5039(1992).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX STRAIN=cv. Marmande;  
 MEDLINE=94071905; PubMed=8250898;  
 RA Roemer S., Hugueney P., Bouvier F., Camara B., Kuntz M.;  
 RT "Expression of the genes encoding the early carotenoid biosynthetic  
 enzymes in Capsicum annum.";  
 RL Biochem. Biophys. Res. Commun. 196:1414-1421(1993).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX STRAIN=cv. Alisa Craig;  
 MEDLINE=88096591; PubMed=3697097;  
 RA Ray J., Bird C.R., Maunders M., Grierson D., Schuch W.;  
 RT "Sequence of PTOM5, a ripening related cDNA from tomato.";  
 RL Nucleic Acids Res. 15:10587-10587(1987).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RX STRAIN=cv. Alisa Craig; TISSUE=leaf;  
 MEDLINE=92322971; PubMed=1623189;  
 RA Ray J., Moureau P., Bird C., Grierson D., Maunders M.,  
 RA Truesdale M., Bramley P., Schuch W.;  
 RT "Cloning and characterization of a gene involved in phytoene  
 synthesis from tomato.";  
 RL Plant Mol. Biol. 19:401-404(1992).  
 CC -1- FUNCTION: CATALYZES THE REACTION FROM PREPHYTOENE DIPHOSPHATE  
 TO PHYTOENE.  
 CC -1- CATALYTIC ACTIVITY: 2 geranylgeranyl diphosphate = diphosphate +  
 prephytoene diphosphate.  
 CC -1- CATALYTIC ACTIVITY: prephytoene diphosphate = diphosphate +  
 phytoene.  
 CC -1- PATHWAY: Carotenoid biosynthesis.  
 CC -1- SUBUNIT: MONOMER (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: Chloroplast.  
 CC -1- DEVELOPMENTAL STAGE: IN SEEDLINGS AND IN LATE STAGES OF FRUIT  
 RIPENING.  
 CC -1- INDUCTION: BY FRUIT RIPENING.  
 CC -1- SIMILARITY: BELONGS TO THE PHYTOENE/SQUALENE SYNTHETASE FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation  
 at the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL: M84744; AAA34153.1; -  
 DR EMBL: X00521; CA688575.1; -  
 DR EMBL: X60441; CA442969.1; -  
 DR EMBL: A21360; CA401348.1; -  
 DR PIR: S06321; S06321.  
 DR PIR: A42102; A42102.  
 DR InterPro: IPR002060; Squ/phyL-synthase.  
 DR Pfam: PF00494; SOS\_PSY; 1  
 DR PROSITE: PS01045; SQUALEN\_PHYTOEN\_SYN\_1; 1.  
 DR PROSITE: PS01045; SQUALEN\_PHYTOEN\_SYN\_2; 1.  
 DR Multifunctional enzyme; Carotenoid biosynthesis; Multigene family;  
 KW Isoprene biosynthesis; Transferase; Chloroplast; Transit peptide.

FT TRANSIT 1 129 CHLOROPLAST (PORENTAL).  
 FT CHAIN 130 412 PHYTOENE SYNTHASE 1.  
 FT CONFLICT 389 407 KLALPLPAVAKSLVPPKRT ->  
 FT CONFLICT 389 405 OVDCTYCIKISICASYKN (IN REF. 3).  
 FT CONFLICT 389 405 KLALPLPAVAKSLVPP ->  
 FT OVDCTYCIKISICASY (IN REF. 4).  
 SO SEQUENCE 412 AA; 46615 MW; DBED05945D6F4C9A CRC64;  
 Query Match 42.68; Score 46; DB 1; Length 412;  
 Best Local Similarity 45.08; Pred. No. 5.2;  
 Matches 9; Conservative 4; Mismatches 3; Indels 4; Gaps 1;  
 QY 1 KMKLFKIGIGKFLHSKRR 20  
 Db 322 KMKLFKIGIGKFLHSKRR 337  
 ID CECA\_HYACE STANDARD; PRT; 64 AA.  
 AC P01507;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 01-APR-1988 (Rel. 07, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Cecropin A precursor (Cecropin C).  
 OS Hyalophora cecropia (Cecropia moth).  
 OC Insecta; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;  
 OC Dityrysta; Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata;  
 OC Hyalophora. Bombycoidea; Saturniidae; Saturniinae; Attacini;  
 ON NCBI\_TaxID=7123;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=91268009; PubMed=1711035;  
 RA Gudmundsson G.H., Lidholm D.-A., Aasling B., Gan R., Boman H.G.;  
 RT "The cecropin locus. Cloning and expression of a gene cluster  
 encoding three antibacterial peptides in Hyalophora cecropia.";  
 RL J. Biol. Chem. 266:11510-11517(1991).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Lidholm D.-A., Gudmundsson G.H., Xanthopoulos K.G., Boman H.G.;  
 RT "Insect immunity: cDNA clones coding for the precursor forms of  
 cecropins A and D, antibacterial proteins from Hyalophora cecropia.";  
 RL FEBS Lett. 226:8-12(1987).  
 RN [3]  
 RP SEQUENCE OF 27-63.  
 RX MEDLINE=83053366; PubMed=7140755;  
 RA Hultmark D., Engstroem A., Bennich H., Kapur R., Boman H.G.;  
 RT "Insect immunity: Isolation and structure of cecropin D and four  
 minor antibacterial components from Cecropia pupae.";  
 RL Eur. J. Biochem. 127:207-217(1982).  
 RN [4]  
 RP PRELIMINARY SEQUENCE OF 27-63.  
 RX MEDLINE=81245158; PubMed=7019715;  
 RA Steiner H., Hultmark D., Engstroem A., Bennich H., Boman H.G.;  
 RT "Sequence and specificity of two antibacterial proteins involved in  
 insect immunity.";  
 RL Nature 292:246-248(1981).  
 RN [5]  
 RP STRUCTURE BY NMR.  
 RX MEDLINE=89088132; PubMed=3207693;  
 RA Holak T.A., Engstroem A., Kraulis P.J., Lindeberg G., Bennich H.,  
 RA Jones T.A., Gronenborn A.M., Clore G.M.;  
 RT "The solution conformation of the antibacterial peptide cecropin A: a  
 nuclear magnetic resonance and dynamical simulated annealing study.";  
 RL Biochemistry 27:7620-7629(1988).  
 CC -1- FUNCTION: CECROPINS HAVE LYTIC AND ANTIBACTERIAL ACTIVITY AGAINST  
 SEVERAL GRAM-POSITIVE AND GRAM-NEGATIVE BACTERIA.  
 CC -1- SUBCELLULAR LOCATION: Secreted.  
 CC -1- PTM: A PROTEIN WITH THE SAME SEQUENCE AS CECROPIN A, BUT LACKING  
 THE CARBOXYL BLOCKING GROUP, WAS ISOLATED AND CALLED CECROPIN C BY  
 REF. 3 AUTHORS.  
 CC -1- SIMILARITY: BELONGS TO THE CECROPIN FAMILY.





DR EMBL: M57952; AAA63288.1; ALT\_FRAME.  
 DR EMBL: S69737; AAC60465.2; ALT\_FRAME.  
 DR TIGR: MT2416; -  
 DR Tuberculin: RY2351C; -  
 KW Hydrolyase; Signal; Membrane; Antigen; Complete proteome.  
 FT SIGNAL 1 37  
 FT CHAIN 38 520  
 FT CONFLICT 327 327 PHOSPHOLIPASE C 1.  
 FT CONFLICT 369 370 T -> N (IN REF. 5).  
 FT CONFLICT 406 407 GE -> AO (IN REF. 5).  
 FT CONFLICT 516 520 RG -> PR (IN REF. 5).  
 FT CONFLICT 516 520 SGLCS -> AGCAAEISR (IN REF. 5).  
 SO SEQUENCE 520 AA; 56138 MW; 5844EF0C064288A7 CRC64;

Query Match 40.7%; Score 44; DB 1; Length 520;  
 Best Local Similarity 33.3%; Pred. No. 14;  
 Matches 5; Conservative 8; Mismatches 2; Indels 0; Gaps 0;

OY 2 WKLFKKGIGKFLHS 16  
 DB 241 WKVYONKGLGRFINT 255

## RESULT 6

CECB\_ANTPE  
 ID CECB\_ANTPE STANDARD; PRT; 35 AA.  
 AC P01509;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 01-APR-1990 (Rel. 14, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Cecropin B.  
 OS Antheraea pernyi (Chinese oak silk moth).  
 OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;  
 OC Insecta; Pelecygota; Neoptera; Endopterygota; Lepidoptera; Glossata;  
 OC Diptera; Bombycoidea; Saturniidae; Saturniinae; Saturniini;  
 OC Antheraea.  
 OC NCBI\_Taxid-7119;  
 RN [1]  
 RP PARTIAL SEQUENCE.  
 RX MEDLINE:83053368; PubMed-6754375;  
 RA Qu X.-M., Steiner H., Engstrom A., Bennich H., Boman H.G.;  
 RT "Insect immunity: Isolation and structure of cecropin B and D from  
 RT pupae of the Chinese oak silk moth, *Antheraea pernyi* L.",  
 RT Eur. J. Biochem. 127:219-224(1982).  
 RN [2]  
 RP SEQUENCE.  
 RX MEDLINE-88108273; PubMed-2962676;  
 RA Craig A.G., Engstrom A., Bennich H., Kamensky I.;  
 RT "Plasma desorption mass spectrometry coupled with conventional  
 RT peptide sequencing techniques.",  
 RT Biomed. Environ. Mass Spectrom. 14:669-673(1987).  
 CC -1- FUNCTION: CECROPINS HAVE LYTIC AND ANTIBACTERIAL ACTIVITY AGAINST  
 CC SEVERAL GRAM-POSITIVE AND GRAM-NEGATIVE BACTERIA.  
 CC -1- SUBCELLULAR LOCATION: Secreted.  
 CC -1- SIMILARITY: BELONGS TO THE CECROPIN FAMILY.  
 CC PIR: A01771; CKA0BP.  
 CC PIR: A54725; A54725.  
 DR InterPro: IPR000875; Cecropin.  
 DR InterPro: IPR003254; IIP-cecropin.  
 DR Pfam: PF00272; cecropin; 1.  
 DR ProDom: PD003996; IIP-cecropin; 1.  
 DR PROSITE: PS00268; CECROPIN; 1.  
 KW Insect immunity; Antibiotic; Hemolymph; Amidation; Multigene family.  
 FT MOD\_RES 35 35 AMIDATION.  
 SO SEQUENCE 35 AA; 3818 MW; 7AFCB0A10E16313 CRC64;

Query Match 39.8%; Score 43; DB 1; Length 35;  
 Best Local Similarity 87.5%; Pred. No. 1.6;  
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 WKLFKKT 8  
 DB 1 WKLFKKT 8

## RESULT 7

YKLL\_CAEEL  
 ID YKLL\_CAEEL STANDARD; PRT; 341 AA.  
 AC P42168;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Putative casein kinase I C03C10.1 in chromosome III (EC 2.7.1.-).  
 GN C03C10.1.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;  
 OC Rhabditidae; Peloderinae; Caenorhabditis.  
 OC NCBI\_Taxid-6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-Bristol N2;  
 RA Berks M.;  
 RL Submitted (AUG-1994) to the EMBL/GenBank/DBJ databases.  
 CC -1- CATALYTIC ACTIVITY: ATP + a protein -> ADP + a phosphoprotein.  
 CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.  
 CC CASEIN KINASE I SUBFAMILY.  
 CC -----

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 -----

DR EMBL: Z35637; CAAB4685.1; -  
 DR HSP: 006486; ICR1.  
 DR WormPep: C03C10.1; CE00872.  
 DR InterPro: IPR00719; Euk\_pkinase.  
 DR InterPro: IPR002290; Ser\_thr\_pkinase.  
 DR Pfam: PF00069; pkinase; 1.  
 DR ProDom: PD000001; Euk\_pkinase; 1.  
 DR PROSITE: PS00107; PROTEIN KINASE ATP; 1.  
 DR PROSITE: PS00108; PROTEIN KINASE ST; 1.  
 DR PROSITE: PS50011; PROTEIN KINASE DOM; 1.  
 KW Hypothetical protein; Transferase; Serine/threonine-protein kinase;  
 KW ATP-binding.  
 FT DOMAIN 16 284  
 FT NP\_BIND 22 30 ATP (BY SIMILARITY).  
 FT BINDING 45 45 ATP (BY SIMILARITY).  
 FT ACT\_SITE 135 135 BY SIMILARITY.  
 SO SEQUENCE 341 AA; 39037 MW; 17AD6E173551BABC CRC64;

Query Match 39.8%; Score 43; DB 1; Length 341;  
 Best Local Similarity 61.5%; Pred. No. 13;  
 Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 1 WKLFKKGIGKFLHS 13  
 DB 15 KYKLIRKIGSGSF 27

## RESULT 8

MTD3\_STYHU  
 ID MTD3\_STYHU STANDARD; PRT; 363 AA.  
 AC Q43138;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE Probable mannitol dehydrogenase 3 (EC 1.1.1.255) (NAD-dependent  
 DE mannitol dehydrogenase 3).  
 GN CAD3.  
 OS Stylosanthes humilis (Towns-ville stylo).

OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids I; Fabales; Fabaceae; Papilionoideae; Aeschynomeneae;  
 OC Stylosanthes.

```

RN NCBI_TaxID=35628;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN=, Paterson; TISSUE=Stem;
RA Nourse J.P., Manners J.M., Curtis M.D., Abrahams S.L., Watson J.M.;
RL submitted (NOV-1994) to the EMBL/GenBank/DBJ databases.
-1- FUNCTION: OXIDIZES MANNITOL TO MANNOSE. PROVIDES THE INITIAL STEP
  BY WHICH TRANSLOCATED MANNITOL IS COMMITTED TO CENTRAL METABOLISM
  AND, BY REGULATING MANNITOL POOL SIZE, IS IMPORTANT IN REGULATING
  SALT TOLERANCE AT THE CELLULAR LEVEL (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: D-mannitol + NAD(+) = D-mannose + NADH.
CC -1- COFACTOR: ZINC (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE ZINC-CONTAINING ALCOHOL DEHYDROGENASE
  FAMILY.
CC -1- CAUTION: WAS ORIGINALLY (REF.1) THOUGHT TO BE A CINNAMYL-ALCOHOL
  DEHYDROGENASE.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL; L36456; AAA7483.1; .
CC DR InterPro: IPR002328; ADH_zinc.
CC DR InterPro: IPR002085; Adh_zn_family.
CC DR Pfam: PF00107; adh_zinc; 1.
CC DR PROSITE; PS00059; ADH_ZINC; 1.
KM Oxidoreductase; NAD; Zinc; Multigene family.
FT METAL 51 51 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 73 73 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 104 104 ZINC (SECOND ATOM) (BY SIMILARITY).
FT METAL 107 107 ZINC (SECOND ATOM) (BY SIMILARITY).
FT METAL 110 110 ZINC (SECOND ATOM) (BY SIMILARITY).
FT METAL 118 118 ZINC (SECOND ATOM) (BY SIMILARITY).
FT METAL 168 168 ZINC (CATALYTIC) (BY SIMILARITY).
SQ SEQUENCE 365 AA; 39551 MW; EBBB217274513D90 CRC64;

Query Match 39.8%; Score 43; DB 1; Length 363;
Best Local Similarity 38.9%; Pred. No. 14;
Matches 7; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

QY 1 KWKLFRKIGIGKFLHSK 18
   1:1 1:1 1:1 1:1
DB 88 KYKGDVGVGVGVESCR 105

RESULT 9
VNUC_INCCA
ID VNUC_INCCA STANDARD; PRT; 565 AA.
AC P08028;
DT 01-AUG-1988 (Rel. 08, Created)
DI 01-AUG-1988 (Rel. 08, Last sequence update)
DR 01-MAY-1992 (Rel. 22, Last annotation update)
DE Nucleoprotein.
GN NP
OS Influenza C virus (strain C/California/78).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenzavirus C.
OX NCBI_TaxID=11555;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=85170522; PubMed=6532006;
RA Nakada S., Creager R.S., Krystal M., Palese P.;
RT "Complete nucleotide sequence of the Influenza C/California/78 virus
  nucleoprotein gene.";
RL Virus Res. 1:433-441(1984).
-1- FUNCTION: ENCAPSIDATES THE NEGATIVE STRAND VIRAL RNA.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

```

```

the European Bioinformatics Institute. There are no restrictions on its
cc use by non-profit institutions as long as its content is in no way
cc modified and this statement is not removed. Usage by and for commercial
cc entities requires a license agreement (See http://www.isb-sib.ch/announce/
cc or send an email to license@isb-sib.ch).
-----
DR EMBL: M17700; AAA43798.1; -
KW Nucleoprotein
NM Sequence 565 AA; 63598 MW; 22C10CB1A85AA84 CRC64;
DQ Query Match 39.8%; Score 43; DB 1; Length 565;
Best Local Similarity 50.0%; Pred. No. 21;
Matches 9; Conservative 2; Mismatches 7; Indels 0; Gaps 0;
OY 1 KWKLFKKIGIKGFLHSK 18
Db 181 KWIEMKHGIGVLADAK 198
RESULT 10
Y870_METUA
ID Y870_METUA STANDARD PRT: 620 AA.
AC Q58280;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein MJ0870.
DN MJ0870.
OS Methanococcus jannaschii.
OC Archaeae; Euryarchaeota; Methanococci; Methanococcales;
OC Methanocaldococcaceae; Methanocaldococcus.
OX NCBI_TaxID=2190;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96337999; PubMed=8688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., Fitzgerald L.M., Claydon R.A., Gocayne J.D.,
RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reisch C.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Georgagen N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Kleink H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT jannaschii."
RL Science 273:1058-1073(1996).
CC -1- SIMILARITY: TO COENZYME F420 HYDROGENASE BETA SUBUNIT.
CC -1- SIMILARITY: TO M.JANASCHII MJ1349, MJ0725 AND MJ0551.
CC -1- SIMILARITY: THE C-TERMINAL DOMAIN IS A 4FE-4S/SIROHEME DOMAIN
CC FOUND IN NITRITE REDUCTASES (EC 1.6.6.4 AND EC 1.7.7.1) AND
CC SULFITE REDUCTASES (EC 1.8.1.2 AND EC 1.8.7.1).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U67531; AAB98876.1; -
DR HSSP: Q45560; LBWE.
DR TIGR: MJ0870; -.
DR InterPro: IPR001450; 4FE4S_ferredoxin.
DR InterPro: IPR000660; Nit_Sir.
DR InterPro: IPR005117; Nit_sir_fer.
DR Pfam: PF00037; fer4_3.
DR Pfam: PF01077; NIR_Sir; 1.
DR Pfam: PF03460; NIR_Sir_ferr; 1.
DR PRINTS: PR00397; SIROHAEM.
DR PROSITE: PS00198; 4FE4S_FERREDOXIN; 2.

```



RT "On the primary structures of lysozyme, cecropins and attacins from  
 RT Hyalophora cecropia.";  
 RL Dev. Comp. Immunol. 9:551-556(1985).  
 RN [4]  
 RP SEQUENCE OF 27-61.  
 RA MEDLINE=61245158; PubMed=7019715;  
 RA Seiler H., Hultmark D., Engstrom A., Bennich H., Boman H.G.;  
 RT "Sequence and specificity of two antibacterial proteins involved in  
 insect immunity.";  
 RT Nature 292:246-248(1981).  
 CC -1- FUNCTION: CECROPINS HAVE LYRIC AND ANTIBACTERIAL ACTIVITY AGAINST  
 CC SEVERAL GRAM-POSITIVE AND GRAM-NEGATIVE BACTERIA.  
 CC -1- SUBCELLULAR LOCATION: Secreted.  
 CC -1- SIMILARITY: BELONGS TO THE CECROPIN FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC DR EMBL; M10309; AAA29184.1; -;  
 CC DR EMBL; X07404; CAA30306.1; -;  
 CC DR EMBL; M34924; AAA29187.1; -;  
 CC DR PIR; A01770; CKMKB.  
 CC DR PIR; S02606; S02606.  
 CC DR InterPro: IPR000875; Cecropin.  
 CC DR InterPro: IPR003254; IIP-cecropin.  
 CC DR Pfam: PF00272; cecropin; 1.  
 CC DR ProDom: PD003996; IIP-cecropin; 1.  
 CC DR PROSITE; PS00268; CECROPIN; 1.  
 CC KW Insect immunity; Antibiotic; Hemolymph; Amidation; Multigene family;  
 CC KW Signal.  
 CC FT SIGNAL 1 22  
 CC FT PROPEP 23 26 REMOVED BY A DIPEPTIDYLPEPTIDASE.  
 CC FT CHAIN 27 61 CECROPIN B.  
 CC FT MOD\_RES 61 61 AMIDATION (G-62 PROVIDE AMIDE GROUP).  
 CC SQ SEQUENCE 62 AA; 6750 MW; 65E309CEB2983B61 CRC64;  
 CC -----  
 CC Query Match 38.9%; Score 42; DB 1; Length 62;  
 CC Best Local Similarity 87.5%; Pred. No. 4;  
 CC Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 CC  
 CC QY 1 KWKLFKTI 8  
 CC Db 27 KWKVFKTI 34  
 CC  
 CC RESULT 13  
 CC CECA\_BOMMO STANDARD; PRT; 63 AA.  
 CC ID CECA\_BOMMO  
 CC AC 027239;  
 CC DT 01-NOV-1997 (Rel. 35, Created)  
 CC DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 CC DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 CC DE Cecropin A precursor.  
 CC GN CECA.  
 CC OS Bombyx mori (Silk moth)  
 CC OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;  
 CC OC Insecta; Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata;  
 CC OC Diptysia; Bombycoidea; Bombycidae; Bombyx.  
 CC OX NCBI\_TaxID=7091;  
 CC RN [1]  
 CC RP SEQUENCE FROM N.A.  
 CC RC STRAIN=C108; TISSUE=Larval fat body;  
 CC RC MEDLINE=94369101; PubMed=7765280;  
 CC RA Yano Y., Matsumoto M., Inoue K., Kawabata T., Morishima I.;  
 RT "Cloning of cDNAs for cecropins A and B, and expression of the genes  
 RT in the silkworm, Bombyx mori.";  
 RT Biosci. Biotechnol. Biochem. 58:1476-1478(1994).  
 CC -1- FUNCTION: CECROPINS HAVE LYRIC AND ANTIBACTERIAL ACTIVITY AGAINST

CC SEVERAL GRAM-POSITIVE AND GRAM-NEGATIVE BACTERIA.  
 CC -1- SUBCELLULAR LOCATION: Secreted.  
 CC -1- TISSUE SPECIFICITY: HIGHEST EXPRESSION IN FAT BODY AND HEMOCYTES.  
 CC IS ALSO EXPRESSED IN MALPIGHIAN TUBE AND TO A MUCH LESSER EXTENT  
 CC IN MIDGUT. NOT PRESENT IN SILK GLAND.  
 CC -1- SIMILARITY: BELONGS TO THE CECROPIN FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC DR EMBL; D17394; BAA04217.1; -;  
 CC DR EMBL; S74376; AAC60515.1; -;  
 CC DR InterPro: IPR000875; Cecropin.  
 CC DR InterPro: IPR003254; IIP-cecropin.  
 CC DR Pfam: PF00272; cecropin; 1.  
 CC DR ProDom: PD003996; IIP-cecropin; 1.  
 CC DR PROSITE; PS00268; CECROPIN; 1.  
 CC KW Insect immunity; Antibiotic; Hemolymph; Amidation; Multigene family;  
 CC KW Signal.  
 CC FT SIGNAL 1 22  
 CC FT PROPEP 23 26 POTENTIAL.  
 CC FT CHAIN 27 61 BY SIMILARITY.  
 CC FT MOD\_RES 61 61 CECROPIN A.  
 CC FT MOD\_RES 61 61 AMIDATION (G-62 PROVIDE AMIDE GROUP)  
 CC (BY SIMILARITY).  
 CC SQ SEQUENCE 63 AA; 6762 MW; 6A1C39975510D86A CRC64;  
 CC -----  
 CC Query Match 38.9%; Score 42; DB 1; Length 63;  
 CC Best Local Similarity 87.5%; Pred. No. 4.1;  
 CC Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 CC  
 CC QY 1 KWKLFKTI 8  
 CC Db 27 KWKVFKTI 34  
 CC  
 CC RESULT 14  
 CC YG31\_HAEIN STANDARD; PRT; 190 AA.  
 CC ID YG31\_HAEIN  
 CC AC P44279;  
 CC DT 01-NOV-1995 (Rel. 32, Created)  
 CC DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 CC DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 CC DE Hypothetical protein H1631 precursor.  
 CC GN H1631.  
 CC OS Haemophilus influenzae.  
 CC OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
 CC OC Haemophilus.  
 CC OX NCBI\_TaxID=727;  
 CC RN [1]  
 CC RP SEQUENCE FROM N.A.  
 CC RC STRAIN=Rd / KW20 / ATCC 51907;  
 CC RC MEDLINE=95350630; PubMed=7542800;  
 CC RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,  
 RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,  
 RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,  
 RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,  
 RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,  
 RA Uterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,  
 RA Fine L.D., Fitchdman J.L., Fuhrmann J.L., Geoghagen N.S.M.,  
 RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,  
 RA Venter J.C.;  
 RT "Whole-genome random sequencing and assembly of Haemophilus influenzae  
 RT Rd.";  
 RT Science 269:496-512(1995).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch))  
 CC -----  
 DR EMBL: U32836; AAC23283.1; -  
 DR TRIGR: H11631; -  
 KW Hypothetical protein; Signal: Complete proteome.  
 FT SIGNAL 1 28  
 FT CHAIN 29 190  
 FT SEQUENCE 190 AA; 22314 MW; E4E3A8C7957EC908 CRC64;  
 SQ  
 Query Match Best Local Similarity 38.9%; Score 42; DB 1; Length 190;  
 Matches 7; Conservative 7; Mismatches 5; Indels 0; Gaps 0;  
 QY 1 KKKLFKKIGIGKFLHSACK 19  
 Db 90 EFRLFYGVSMGEFLESIDK 108  
 RESULT 15  
 KCL1\_BOVIN STANDARD; PRT; 325 AA.  
 AC P35506;  
 DT 01-JUN-1994 (Rel. 29, Created)  
 DT 01-JUN-1994 (Rel. 29, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Casein kinase I, alpha isoform (EC 2.7.1.-) (CKI-alpha) (CKI).  
 GN CSNK1A1.  
 OS Bos taurus (Bovine).  
 OS Oryctolagus cuniculus (Rabbit), and  
 OS Ovis aries (Sheep).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
 OC Bovidae; Bovinae; Bos.  
 OX NCBI\_TaxID=9913, 9986, 9940;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
 RC SPECIES-Bovine; TISSUE-Brain;  
 RX MEDLINE=92052129; PubMed=1946367;  
 RA Rowles J., Slaughter C., Moomaw C., Hsu J., Cobb M.H.;  
 RT Purification of casein kinase I and isolation of cDNAs encoding  
 RT multiple casein kinase I-like enzymes.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 88:9548-9552(1991).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES-Rabbit; TISSUE-Skeletal muscle;  
 RA Zhai L., Depoell-Roach A.A., Roach P.J.;  
 RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES-Sheep; TISSUE-Blood;  
 RA Yamamoto N.;  
 RT Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: CASEIN KINASES ARE OPERATIONAL/DEFINED BY THEIR  
 CC PREFERENTIAL UTILIZATION OF ACIDIC PROTEINS SUCH AS CASEINS  
 CC AS SUBSTRATES. IT CAN PHOSPHORYLATE A LARGE NUMBER OF PROTEINS.  
 CC -1- SUBUNIT: MONOMER.  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.  
 CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.  
 CC CASEIN KINASE I SUBFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch))  
 CC -----  
 DR EMBL: M76543; AAA30451.1; -

DR EMBL: U59166; AAE03992.1; -  
 DR EMBL: AB050945; BAB17769.1; -  
 DR HSSP: Q06486; 1CKI.  
 DR InterPro: IPR000719; Euk\_Pkinase.  
 DR InterPro: IPR002290; Ser\_thr\_Pkinase.  
 DR Pfam: PF00069; Pkinase; 1.  
 DR ProDom: PD000001; Euk\_Pkinase; 1.  
 DR PROSITE: PS00107; PROTEIN\_KINASE\_ATP; 1.  
 DR PROSITE: PS00108; PROTEIN\_KINASE\_ST; 1.  
 DR PROSITE: PS50011; PROTEIN\_KINASE\_DOM; 1.  
 KW Transferase; Serine/threonine-protein kinase; ATP-binding;  
 KW Multigene family.  
 FT DOMAIN 17 285  
 FT NP\_BIND 23 31  
 FT BINDING 46 46  
 FT ACT\_SITE 136 136  
 FT SEQUENCE 325 AA; 37567 MW; B84DC84BDDC17854 CRC64;  
 Query Match Best Local Similarity 38.9%; Score 42; DB 1; Length 325;  
 Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;  
 QY 1 KKKLFKKIGIGKFLHSACK 13  
 Db 16 KKKLVKRGSGSF 28  
 Search completed: June 6, 2003, 13:23:58  
 Job time : 8 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 6, 2003, 13:21:26 ; Search time 23.5 Seconds  
(without alignments)  
175.359 Million cell updates/sec

Title: US-10-081-418-1  
Perfect score: 108  
Sequence: 1 KMKLFKKIGKFLHSAKF 20

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 671580 segs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

SPTREMBL\_21:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mmc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_prodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_virus:\*  
16: sp\_bacteriaph:\*  
17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	48	44.4	309	16	Q8RC36 thermomaneer
2	47.5	44.0	237	17	Q9U283
3	47.5	44.0	427	16	Q9K5P7
4	47.5	44.0	501	3	Q96U53
5	47	43.5	1562	3	Q60U34
6	46	42.6	376	10	Q8VYE2
7	46	42.6	414	10	Q9LG22
8	46	42.6	432	10	Q40166
9	45	42.6	666	16	Q98OC7
10	45	41.7	88	2	Q9AINE
11	45	41.7	154	11	Q9CYL5
12	45	41.7	319	5	Q9U2N8
13	45	41.7	925	16	Q8YBF0
14	44	40.7	205	16	Q99TD7
15	44	40.7	353	3	Q08980
16	44	40.7	499	5	Q9VN08

17	43	39.8	170	16	Q9RX21	Q9rx21 delnoccocus
18	43	39.8	199	12	Q9IO51	Q9io51 influenzavi
19	43	39.8	199	12	Q9IO50	Q9io50 influenzavi
20	43	39.8	199	12	Q9O217	Q9o217 influenzavi
21	43	39.8	199	12	Q9O2P4	Q9o2p4 influenzavi
22	43	39.8	199	12	Q9O205	Q9o205 influenzavi
23	43	39.8	199	12	Q9O203	Q9o203 influenzavi
24	43	39.8	199	12	Q9O201	Q9o201 influenzavi
25	43	39.8	199	12	Q9O1Z9	Q9o1z9 influenzavi
26	43	39.8	199	12	Q9O1Z7	Q9o1z7 influenzavi
27	43	39.8	199	12	Q9O1Z5	Q9o1z5 influenzavi
28	43	39.8	199	12	Q9O1Z3	Q9o1z3 influenzavi
29	43	39.8	199	12	Q9O1Z1	Q9o1z1 influenzavi
30	43	39.8	199	12	Q9O1Y9	Q9o1y9 influenzavi
31	43	39.8	199	12	Q9O1Y7	Q9o1y7 influenzavi
32	43	39.8	199	12	Q8QZ11	Q8qz11 influenzavi
33	43	39.8	199	12	Q8QZ10	Q8qz10 influenzavi
34	43	39.8	199	12	Q8QZ09	Q8qz09 influenzavi
35	43	39.8	199	12	Q8QZ08	Q8qz08 influenzavi
36	43	39.8	199	12	Q8QZ07	Q8qz07 influenzavi
37	43	39.8	199	12	Q8QZ06	Q8qz06 influenzavi
38	43	39.8	199	12	Q8QZ05	Q8qz05 influenzavi
39	43	39.8	199	12	Q8QZ04	Q8qz04 influenzavi
40	43	39.8	199	12	Q8QZ03	Q8qz03 influenzavi
41	43	39.8	199	12	Q8QZ02	Q8qz02 influenzavi
42	43	39.8	199	12	Q8QZ01	Q8qz01 influenzavi
43	43	39.8	199	12	Q8QZ00	Q8qz00 influenzavi
44	43	39.8	199	12	Q8QZ00	Q8qz00 influenzavi
45	43	39.8	199	12	Q8QZ00	Q8qz00 influenzavi

## ALIGNMENTS

RESULT 1	
ID Q8RC36	PRELIMINARY; PRT; 309 AA.
AC Q8RC36;	
DT 01-JUN-2002 (TREMBLrel. 21, Created)	
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)	
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)	
DE ABC-type dipeptide/oligopeptide/nickel transport systems, permease components.	
GN DPPP4 OR TTE0612.	
OS Thermoanaerobacter tengcongensis.	
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;	
OC Thermoanaerobacteriales; Thermoanaerobacteriaceae; Thermoanaerobacter.	
OX NCBI_TaxID=119072;	
RN [1]	
RP SEQUENCE FROM N.A.	
RC STRAIN=MB4T / JCM11007;	
RX MEDLINE=21992816; PubMed=11997336;	
RA Bao Q., Tian Y., Li W., Xu Z., Xuan Z., Hu S., Dong W., Yang J.,	
RA Chen Y., Xue Y., Xu Y., Lai X., Huang L., Dong X., Ma Y., Ling L.,	
RA Tan H., Chen R., Wang J., Yu J., Yang H.;	
RT "A complete sequence of T. tengcongensis genome.";	
RL Genome Res. 12:689-700(2002).	
DR EMBL; AE013030; AAM23882.1; -	
KW Complete proteome.	
SQ SEQUENCE 309 AA; 34754 MW; A90E0686091EFF74 CRC64;	
QY	1 KMKLFKKIGKFLHS 16
DB	157 KMKLFPPVSGWGTFAHT 172
Matches	9; Conservative 44.4%; Score 48; DB 16; Length 309;
	Best Local Similarity 56.2%; Pred. No. 11;
	Mismatches 6; Indels 0; Gaps 0;
RESULT 2	
ID Q9U283	PRELIMINARY; PRT; 237 AA.

AC Q9U283;  
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
 DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)  
 DE Arginase related (EC 3.5.3.1).  
 GN PAB1359.  
 OS Pyrococcus abyssi.  
 OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;  
 OC Pyrococcus.  
 OX NCBI\_TaxID=29292;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ORSAY;  
 RA Heilig R.;  
 RT "Pyrococcus abyssi genome sequence: insights into archaeal chromosome structure and evolution."  
 RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AJ248287; CAB50176.1; -.  
 DR InterPro: IPR000287; Arginase.  
 DR Pfam: PF00491; arginase; 1.  
 DR PROSITE: PS00148; ARGINASE\_2; UNKNOWN\_1.  
 KW Complete proteome.  
 SQ SEQUENCE 237 AA; 27191 MW; E044EF7EB136463D CRC64;

Query Match 44.0%; Score 47.5; DB 17; Length 237;  
 Best Local Similarity 39.1%; Pred. No. 9.9;  
 Matches 9; Conservative 6; Mismatches 5; Indels 3; Gaps 1;

OY 1 KWLF---KKGIGKFLHSAKKF 20  
 DB 135 KMKPFWKRGVGRKATKLVETAKK 157

## RESULT 3

O9K5P7 PRELIMINARY; PRT; 427 AA.  
 AC O9K5P7;  
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
 DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)  
 DE Hypothetical protein BH4041.  
 GN BH4041.  
 OS Bacillus halodurans.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;  
 OC Bacillaceae; Bacillus.  
 OX NCBI\_TaxID=86665;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C-125 / JCM 9153;  
 RX MEDLINE=20512582; PubMed=11058132;  
 RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,  
 RA Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kohara S.,  
 RA Horikoshi K.;  
 RT "Complete genome sequence of the alkaliphilic bacterium Bacillus  
 RT halodurans and genomic sequence comparison with Bacillus subtilis."  
 RL Nucleic Acids Res. 28:4317-4331(2000).  
 DR EMBL: AP001520; BAB07760.1; -.  
 DR InterPro: IPR003346; Transposase\_20.  
 DR InterPro: IPR002525; Transposase\_9.  
 DR Pfam: PF02371; Transposase\_20; 1.  
 DR Pfam: PF01548; Transposase\_9; 1.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 427 AA; 48695 MW; F7E8099BCDA45C5C CRC64;

Query Match 44.0%; Score 47.5; DB 16; Length 427;  
 Best Local Similarity 45.8%; Pred. No. 18;  
 Matches 11; Conservative 3; Mismatches 5; Indels 5; Gaps 1;

OY 1 KMKLFKIGIG-----KFLHSARK 19  
 DB 217 KMKPFWKRGVGRKATKLVETAKK 240

## RESULT 4

O96U53 PRELIMINARY; PRT; 501 AA.  
 AC O96U53;  
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
 DE Hypothetical 56.9 kDa protein.  
 GN B24P7.050.  
 OS Neurospora crassa.  
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
 OC Sordariales; Sordariaceae; Neurospora.  
 OX NCBI\_TaxID=5141;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Schulte U., Aign V., Hohnsels J., Brandt P., Fartmann B., Holland R.,  
 RA Nyakatura G., Mewes H.W., Mannhaupt G.;  
 RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA German Neurospora genome project;  
 RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AL389890; CAD11797.1; -.  
 DR InterPro: IPR000822; ZnF\_C2H2.  
 DR SMART: SM00355; ZnF\_C2H2; 3.  
 DR PROSITE: PS00028; ZINC\_FINGER\_C2H2\_1; UNKNOWN\_2.  
 KW DNA-binding; Hypothetical protein; Zinc-finger.  
 SQ SEQUENCE 501 AA; 56866 MW; 5B7D09A2ED6C90E9 CRC64;

Query Match 44.0%; Score 47.5; DB 3; Length 501;  
 Best Local Similarity 50.0%; Pred. No. 21;  
 Matches 10; Conservative 4; Mismatches 5; Indels 1; Gaps 1;

OY 2 WKLFK-KGIGKFLHSAKKF 20  
 DB 320 WTLMKPSCGCHKYHMSAKKF 339

## RESULT 5

O60034 PRELIMINARY; PRT; 1562 AA.  
 AC O60034;  
 DT 01-AUG-1998 (TrEMBLrel. 07, Created)  
 DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
 DE ATP-binding multidrug cassette transporter.  
 OS Botrytis cinerea (Botryotinia fuckeliana).  
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Leotiomycetes;  
 OC Helotiales; Sclerotiniaceae; Botryotinia.  
 OX NCBI\_TaxID=40559;  
 RN [1]  
 RP SEQUENCE OF 1-1283 FROM N.A.  
 RC STRAIN=SAS56;  
 RX MEDLINE=97324590; PubMed=9180695;  
 RA Del Sorbo G., Van Nistelrooy H., Andrade A., van Kan J.A.L., Balzi,  
 RA De Waard M.;  
 RT "Multidrug resistance in Aspergillus nidulans involves novel ATP-  
 RT binding cassette transporters."  
 RL Mol. Gen. Genet. 254:417-426(1997).  
 CC 1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY.  
 DR EMBL: Z68906; CA93142.1; -.  
 DR InterPro: IPR003593; AAA\_ATPase.  
 DR InterPro: IPR003439; ABC\_transporter.  
 DR InterPro: IPR000577; FGGY\_kin.  
 DR InterPro: IPR001092; HLH\_basic.  
 DR Pfam: PF00005; ABC\_tran; 2.



DR ProDom; PD000006; ABC\_transport; 2.  
 DR SMART; SM00382; AAA; 1.  
 DR PROSITE; PS00211; ABC\_TRANSPORTER; 1.  
 DR PROSITE; PS00445; PGCG\_KINASES; 2; UNKNOWN\_1.  
 DR PROSITE; PS00038; HELIX\_LOOP\_HELIX; UNKNOWN\_1.  
 KW ATP-binding; transport.  
 SQ SEQUENCE 1562 AA; 173535 MW; 89036ECCE3D24919 CRC64;

Query Match 43.5%; Score 47; DB 3; Length 1562;  
 Best Local Similarity 30.8%; Pred. No. 82;  
 Matches 8; Conservative 7; Mismatches 3; Indels 8; Gaps 1;

OY 1 KWK-----LFKKIGIGKFLHSAKF 18  
 DB 1497 KWRNFGIFLGILYQQLGVGLFLHQOR.1522

RESULT 6  
 OYVE2 PRELIMINARY; PRT; 376 AA.  
 ID OYVE2;  
 DT 01-MAR-2002 (TREMBLrel. 20, Created)  
 DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE Hypothetical 42.6 kDa protein.  
 GN AtG55900.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC Eurosid II; Brassicales; Brassicaceae; Arabidopsids.  
 RX NCBI\_TaxID=3702;  
 RP SEQUENCE FROM N.A.

RA Yamada K., Banh J., Chan M.M., Chang C.H., Chang E., Dale J.M.,  
 RA Deng J.M., Goldsmith A.D., Lee J.M., Onodera C.S., Quach H.L.,  
 RA Tang C., Toriumi M., Wu H.C., Yamamura Y., Yu G., Bowser L.,  
 RA Carinanci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T.,  
 RA Kamiya A., Karlin-Neumann G., Kawai J., Kim C., Lam B., Lin J.,  
 RA Meyers M.C., Miranda M., Narusaka M., Nguyen M., Palm C.J.,  
 RA Sakurai T., Satou M., Seki J.R., Shinozaki A., Shinozaki K.,  
 RA Davis R.W., Ecker J.R., Theologis A.,  
 RT "Full Length cDNA of gene AtG55900 (GI:15222797)."  
 RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AY072140; AAL59962.1;  
 DR InterPro: IPR004274; NIF.  
 DR Pfam: PF03031; NIF; 1.  
 KW Hypothetical protein.  
 SQ SEQUENCE 376 AA; 42639 MW; 160466DD597215AA CRC64;

Query Match 42.6%; Score 46; DB 10; Length 376;  
 Best Local Similarity 42.1%; Pred. No. 28;  
 Matches 8; Conservative 3; Mismatches 8; Indels 0; Gaps 0;

OY 2 WKLFFKIGIGKFLHSAKF 20  
 DB 212 WRTFKRPGVDALFHLGKF 230

RESULT 7  
 OYLG22 PRELIMINARY; PRT; 414 AA.  
 ID OYLG22;  
 AC OYLG22;  
 DT 01-OCT-2000 (TREMBLrel. 15, Created)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE F14316.15.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC Eurosid II; Brassicales; Brassicaceae; Arabidopsids.  
 RX NCBI\_TaxID=3702;  
 RP SEQUENCE FROM N.A.

RA Shinn P., Brooks S., Buehler E., Chao O., Johnson-Hopson C., Khan S.,  
 RA Kim C., Altafi H., Bel O., Chin C., Chou J., Choi E., Conn L.,  
 RA Conway A., Gonzalez A., Hansen N., Howling B., Koo T., Lam B., Lee J.,  
 RA Lanz C., Li J., Liu A., Liu K., Liu S., Mukharasy N., Nguyen M.,  
 RA Palm C., Pham P., Sakano H., Schwartz J., Southwick A., Thaverl A.,  
 RA Toriumi M., Vaysberg M., Yu G., Federspiel N.A., Theologis A.,  
 RA Ecker J.R.,  
 RT "Genomic sequence for Arabidopsis thaliana BAC F14316 from chromosome  
 RT I.";  
 RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Ecker J.R.,  
 RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA Ecker J.R.,  
 RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RA Ecker J.R.,  
 RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RA Ecker J.R.,  
 RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.  
 RN [6]  
 RP SEQUENCE FROM N.A.  
 RA Khan S., Kim C., Altafi H., Bel O., Chin C., Chou J., Choi E.,  
 RA Conn L., Conway A., Gonzalez A., Hansen N., Howling B., Koo T., Lam B.,  
 RA Lee J., Lenz C., Li J., Liu A., Liu K., Liu S., Mukharasy N.,  
 RA Nguyen M., Palm C., Pham P., Sakano H., Schwartz J., Southwick A.,  
 RA Thaverl A., Toriumi M., Vaysberg M., Yu G., Davis R., Federspiel N.,  
 RA Theologis A., Ecker J.,  
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AC002304; AAF79316.1;  
 DR InterPro: IPR004274; NIF.  
 DR Pfam: PF03031; NIF; 1.  
 SQ SEQUENCE 414 AA; 46605 MW; 1D21A52D01F8D413 CRC64;

Query Match 42.6%; Score 46; DB 10; Length 414;  
 Best Local Similarity 42.1%; Pred. No. 31;  
 Matches 8; Conservative 3; Mismatches 8; Indels 0; Gaps 0;

OY 2 WKLFFKIGIGKFLHSAKF 20  
 DB 212 WRTFKRPGVDALFHLGKF 230

RESULT 8  
 OY0166 PRELIMINARY; PRT; 432 AA.  
 ID OY0166;  
 AC OY0166;  
 DT 01-NOV-1996 (TREMBLrel. 01, Created)  
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)  
 DE Mutant phytoene synthase.  
 GN GTOM5.  
 OS Lycopersicon esculentum (Tomato).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 OC Asteridae; euasterids I; Solanales; Solanaceae; Solanum.  
 RX NCBI\_TaxID=4081;  
 RP SEQUENCE FROM N.A.  
 RA STRAIN=ATISA CRAIG; TISSUE=PERICARP;  
 RX MEDLINE=93344508; PubMed=8343597;  
 RA Fray R.G., Giersen D.,  
 RT "Identification and genetic analysis of normal and mutant phytoene  
 RT synthase genes by of tomato by sequencing, complementation and co-  
 RT suppression.";  
 RL Plant Mol. Biol. 22:589-602(1993).

Db 109 WKNTKKIGVGK 119

```

RESULT 12
ID Q9U2N8 PRELIMINARY; PRT; 319 AA.
AC Q9U2N8;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-OCT-2001 (TREMBLrel. 18, Last annotation update)
DE Y37H2A.8 protein.
GN Y37H2A.8.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C.elegans: A platform for
RT investigating biology.";
RL Science 282:2012-2018(1998).
DR EMBL; AL132863; CAB60572.1;
DR InterPro; IPR002900; DUF38.
DR InterPro; IPR001810; F-box.
DR Pfam; PF01827; DUF38; 1.
DR Pfam; PF00646; F-box; 1.
SQ SEQUENCE 319 AA; 37863 MW; 69C3C1E2CEA8FD14 CRC64;

Query Match 41.7%; Score 45; DB 5; Length 319;
Best Local Similarity 46.7%; Pred. No. 35;
Matches 7; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

OY 1 KMKLFKKIGIGKFLH 15
Db 249 EMKSKDTKLGKFFH 263

RESULT 13
ID Q8YBF0 PRELIMINARY; PRT; 925 AA.
AC Q8YBF0;
DT 01-MAR-2002 (TREMBLrel. 20, Created)
DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Nitrate reductase alpha chain (BC 1.7.99.4).
GN BME110950.
OS Brucella melitensis.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Brucellaceae; Brucella.
OX NCBI_TaxID=29459;
RN [1]
RP SEQUENCE FROM N.A.
RL STRAIN=16M / ATCC 23456 / BIOTYPE 1;
RX MEDLINE=20020109; PubMed=11756688;
RA DelVecchio V.G., Kapral V., Redkar R.J., Patra G., Mijer C., Los T.,
RA Ivanova N., Anderson I., Bhattacharya A., Lykidis A., Reznik G.,
RA Jabloner L., Larsen N., D'Souza M., Bernal A., Goltsman E.,
RA Selkov E., Elzer P.H., Hagius S., O'Callaghan D., Letesson J.-J.,
RA "The genome sequence of the facultative intracellular pathogen
RT Brucella melitensis.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:443-448(2002).
DR EMBL; AE009728; AML54192.1;
DR InterPro; IPR001467; Prok_Moxred.
DR Pfam; PF00384; molybdopterin; 1.
DR Pfam; PF01568; molybdopterin; 1.
DR PROSITE; PS00490; MOLYBDOPTERIN_PROK_2; 1.
DR PROSITE; PS00932; MOLYBDOPTERIN_PROK_3; 1.
DR Oxidoreductase; Complete proteome.
SQ SEQUENCE 925 AA; 102574 MW; 15278350CB06033F CRC64;

```

```

Query Match 41.7%; Score 45; DB 16; Length 925;
Best Local Similarity 50.0%; Pred. No. 1e+02;
Matches 10; Conservative 2; Mismatches 8; Indels 0; Gaps 0;

OY 1 KMKLFKKIGIGKFLHSAKF 20
Db 72 KMNLEKDGKQVLELAKSF 91

RESULT 14
ID Q99TD7 PRELIMINARY; PRT; 205 AA.
AC Q99TD7;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Hypothetical protein SAV1727.
GN SAV1727 OR SA1548.
OS Staphylococcus aureus (strain Mu50 / ATCC 700699), and
OS Staphylococcus aureus (strain N315).
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Staphylococcus.
OX NCBI_TaxID=158878, 158879;
RN [1]
RP SEQUENCE FROM N.A.
RL SPECIES=S.aureus (strain Mu50), and S.aureus (strain N315);
RX MEDLINE=21311952; PubMed=11418146;
RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,
RA Cul L., Oguchi A., Aoki K.-I., Nagai Y., Iino T.,
RA Kanamori M., Matsumaru H., Maruyama A., Murakami H., Hosoyama A.,
RA Mizutani T., Takahashi N.K., Sawano T., Inoue R.-I., Kaito C.,
RA Sekimizu K., Hirakawa H., Kohara S., Goto S., Yabuzaki J.,
RA Kanehisa M., Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T.,
RA Hattori M., Ogasawara N., Hayashi H., Hiramatsu K.;
RT "Whole genome sequencing of methicillin-resistant Staphylococcus
RT aureus.";
RL Lancet 357:1225-1240(2001).
DR EMBL; AP003363; BAB57889.1;
DR EMBL; AP003134; BAB42816.1;
DR InterPro; IPR004552; AGP_acyltrn.
DR Pfam; PF01553; AGP_acyltransferase; 1.
DR TIGRFAMS; TIGR00530; AGP_acyltrn; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 205 AA; 23074 MW; 0078ADCE660218E CRC64;

Query Match 40.7%; Score 44; DB 16; Length 205;
Best Local Similarity 55.0%; Pred. No. 32;
Matches 11; Conservative 1; Mismatches 8; Indels 0; Gaps 0;

OY 1 KMKLFKKIGIGKFLHSAKF 20
Db 63 KKELFKKWIGKFLTLNLF 82

RESULT 15
ID Q08980 PRELIMINARY; PRT; 353 AA.
AC Q08980;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Chromosome XVI reading frame ORF YPL264C.
GN YPL264C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RL Delius H., Hebling U.;
RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.

```

RN [2]  
RP SEQUENCE FROM N.A.  
RA Dueterhoeft A., Floeth M., Fritz M., Hilbert H., Moestl D.;  
RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RA MIPS;  
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.  
DR EMBL: Z73620; CA97999.1; -.  
DR SGD: S0006185; YPL264C.  
DR InterPro: IPR000620; DUF6.  
DR Pfam: PF00892; DUF6; 2.  
SQ SEQUENCE 353 AA; 39702 MW; 68A971BC5DB92988 CRC64;  
  
Query Match 40.7%; Score 44; DB 3; Length 353;  
Best Local Similarity 46.7%; Pred. No. 56;  
Matches 7; Conservative 3; Mismatches 5; Indels 0; Gaps 0;  
  
QY 1 KWKLFRKIGIGKFLH 15  
: | | | | | | | | | |  
Db 253 QWGLFLNLGIGSGFIH 267

Search completed: June 6, 2003, 13:24:53  
Job time : 24.5 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 6, 2003, 13:22:41 ; Search time 15.5 seconds

(Without alignments)  
133,213 Million cell updates/sec

Title: US-10-081-418-1

Perfect score: 108

Sequence: 1 KWKLFKRGIGKFLHSAKRF 20

Scoring table: BLOSUM62

Searched: 392085 seqs, 103240269 residues

Total number of hits satisfying chosen parameters: 392085

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Testing first 45 summaries

Database : Published Applications AA:\*

- 1: /cgn2\_6/ptodata/2/pubppa/US08\_NEW\_PUB.pep.\*
- 2: /cgn2\_6/ptodata/2/pubppa/US08\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/ptodata/2/pubppa/US06\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/ptodata/2/pubppa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/2/pubppa/US07\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/ptodata/2/pubppa/US07\_PUBCOMB.pep.\*
- 7: /cgn2\_6/ptodata/2/pubppa/PC08\_PUBCOMB.pep.\*
- 8: /cgn2\_6/ptodata/2/pubppa/US08\_PUBCOMB.pep.\*
- 9: /cgn2\_6/ptodata/2/pubppa/US09\_NEW\_PUB.pep.\*
- 10: /cgn2\_6/ptodata/2/pubppa/US09\_PUBCOMB.pep.\*
- 11: /cgn2\_6/ptodata/2/pubppa/US10\_NEW\_PUB.pep.\*
- 12: /cgn2\_6/ptodata/2/pubppa/US10\_PUBCOMB.pep.\*
- 13: /cgn2\_6/ptodata/2/pubppa/US60\_NEW\_PUB.pep.\*
- 14: /cgn2\_6/ptodata/2/pubppa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	108	100.0	20	US-10-081-418-1	Sequence 1, Appli
2	64	59.3	32	US-09-908-139-6	Sequence 6, Appli
3	63	58.3	23	US-09-820-053A-7	Sequence 7, Appli
4	63	58.3	23	US-09-904-753-2	Sequence 2, Appli
5	63	58.3	23	US-09-030-619-211	Sequence 21, App
6	63	58.3	26	US-09-904-753-13	Sequence 13, Appli
7	63	58.3	26	US-09-030-619-120	Sequence 120, App
8	63	58.3	28	US-09-908-139-1	Sequence 1, Appli
9	63	58.3	28	US-09-909-652-7	Sequence 121, App
10	63	58.3	28	US-09-030-619-121	Sequence 121, App
11	63	58.3	303	US-09-917-340-4	Sequence 4, Appli
12	60	55.6	23	US-09-820-053A-146	Sequence 146, App
13	59	54.6	33	US-09-908-139-9	Sequence 9, Appli
14	56	51.9	23	US-09-904-753-1	Sequence 1, Appli
15	56	51.9	23	US-09-030-619-210	Sequence 210, App
16	54	50.0	22	US-09-807-720-3	Sequence 3, Appli
17	51	47.2	15	US-09-820-053A-33	Sequence 33, Appli
18	50	46.3	22	US-09-820-053A-24	Sequence 24, Appli
19	50	46.3	22	US-09-904-753-3	Sequence 3, Appli

20	50	46.3	22	9	US-09-904-753-4	Sequence 4, Appli
21	49.5	45.8	31	9	US-09-908-139-8	Sequence 8, Appli
22	48.5	44.9	23	9	US-09-908-139-12	Sequence 12, Appli
23	48	44.4	30	9	US-09-908-139-7	Sequence 7, Appli
24	48	44.4	376	9	US-09-738-626-6975	Sequence 6975, Ap
25	46.5	43.1	30	9	US-09-908-139-11	Sequence 11, Appli
26	46	42.6	412	10	US-09-371-307-76	Sequence 76, Appli
27	45	41.7	37	10	US-09-904-753-5	Sequence 5, Appli
28	45	41.7	37	10	US-09-030-619-176	Sequence 176, App
29	45	41.7	535	9	US-10-020-079-16	Sequence 16, Appli
30	45	41.7	548	9	US-10-020-079-14	Sequence 14, Appli
31	45	41.7	648	9	US-10-020-079-12	Sequence 12, Appli
32	45	41.7	654	9	US-10-020-079-10	Sequence 10, Appli
33	45	41.7	751	9	US-10-020-079-8	Sequence 8, Appli
34	45	41.7	764	9	US-10-020-079-6	Sequence 6, Appli
35	45	41.7	776	9	US-10-020-079-4	Sequence 4, Appli
36	45	41.7	789	9	US-10-020-079-22	Sequence 22, Appli
37	45	41.7	838	9	US-10-020-079-40	Sequence 40, Appli
38	45	41.7	851	9	US-10-020-079-38	Sequence 38, Appli
39	45	41.7	863	9	US-10-020-079-32	Sequence 32, Appli
40	45	41.7	864	9	US-10-020-079-30	Sequence 30, Appli
41	45	41.7	870	9	US-10-020-079-28	Sequence 28, Appli
42	45	41.7	876	9	US-10-020-079-20	Sequence 20, Appli
43	45	41.7	889	9	US-10-020-079-18	Sequence 18, Appli
44	45	41.7	895	9	US-10-020-079-16	Sequence 16, Appli
45	45	41.7	951	9	US-10-020-079-36	Sequence 36, Appli

## ALIGNMENTS

RESULT 1  
US-10-081-418-1  
Sequence 1, Application US/10081418,  
Publication No. US20030096745A1  
GENERAL INFORMATION:  
APPLICANT: HAHM, Kyung-Soo  
APPLICANT: LEE, Yong Gun  
APPLICANT: PARK, Doonkyung  
TITLE OF INVENTION: No. US20030096745A1 peptides with increased + charge and hyd.  
TITLE OF INVENTION: substituting one or more amino acids of CA-MA peptide and  
FILE REFERENCE: 428.1014  
CURRENT APPLICATION NUMBER: US/10/081,418  
NUMBER OF SEQ ID NOS: 2  
SOFTWARE: Kopatentlin 1.71  
SEQ ID NO 1  
LENGTH: 20  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: CA-MA peptide made by fusing 1-8 amino acid of secretin A and  
US-10-081-418-1

Query Match  
Best local similarity 100.0%; Score 108; DB 9; Length 20;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 KWKLFKRGIGKFLHSAKRF 20  
 Db 1 KWKLFKRGIGKFLHSAKRF 20  
 US-09-908-139-6  
 ; Sequence 6, Application US/09908139  
 ; Publication No. US20030096949A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Hancock, Robert E. W.  
 ; APPLICANT: Gough, Monisha A.

APPLICANT: Patrzykat, Aleksander  
APPLICANT: Woods, Donald  
APPLICANT: Jia, Xiaoyan  
TITLE OF INVENTION: ANTI-ENDOTOXIC, ANTIMICROBIAL CATIONIC  
FILE REFERENCE: 07422/016001  
CURRENT APPLICATION NUMBER: US/09/908,139  
CURRENT FILING DATE: 2001-07-17  
PRIOR APPLICATION NUMBER: 09/143,124  
PRIOR FILING DATE: 1998-08-28  
NUMBER OF SEQ ID NOS: 25  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 6  
LENGTH: 32  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: antimicrobial cationic peptide  
US-09-908-139-6

Query Match 59.3%; Score 64; DB 9; Length 32;  
Best Local Similarity 72.2%; Pred. No. 0.0017;  
Matches 13; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 1 KWKLFKKIGIGKFLHSAK 18  
DB 1 KWKLFKKIGIGAVLKVK 18

RESULT 3  
US-09-820-053A-7  
Sequence 7, Application US/09820053A  
Publication No. US20030083243A1  
GENERAL INFORMATION:  
APPLICANT: Owen, Donald R.  
TITLE OF INVENTION: SHORT BIOACTIVE PEPTIDES  
FILE REFERENCE: HELX027  
CURRENT APPLICATION NUMBER: US/09/820,053A  
CURRENT FILING DATE: 2001-03-28  
NUMBER OF SEQ ID NOS: 165  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 7  
LENGTH: 23  
TYPE: PRT  
ORGANISM: ARTIFICIAL SEQUENCE  
FEATURE:  
OTHER INFORMATION: SYNTHETIC SEQUENCE  
NAME/KEY: MOD\_RES  
LOCATION: (23)  
OTHER INFORMATION: AMIDATION  
US-09-820-053A-7

Query Match 58.3%; Score 63; DB 9; Length 23;  
Best Local Similarity 100.0%; Pred. No. 0.0018;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 9 GIGKFLHSAKFF 20  
DB 1 GIGKFLHSAKFF 12

RESULT 4  
US-09-904-753-2  
Sequence 2, Application US/09904753  
Publication No. US20030092612A1  
GENERAL INFORMATION:  
APPLICANT: Lynos, Robert T  
TITLE OF INVENTION: Use of Antimicrobial Peptides as Preservatives in  
TITLE OF INVENTION: Ophthalmic Preparations, Including Solutions,  
FILE REFERENCE: 2973 ver 2  
CURRENT APPLICATION NUMBER: US/09/904,753  
CURRENT FILING DATE: 2001-07-13

PRIOR APPLICATION NUMBER: WO 96/25183  
PRIOR FILING DATE: 1996-08-22  
NUMBER OF SEQ ID NOS: 14  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 2  
LENGTH: 23  
TYPE: PRT  
ORGANISM: Xenopus laevis  
US-09-904-753-2

Query Match 58.3%; Score 63; DB 9; Length 23;  
Best Local Similarity 100.0%; Pred. No. 0.0018;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 9 GIGKFLHSAKFF 20  
DB 1 GIGKFLHSAKFF 12

RESULT 5  
US-09-030-619-211  
Sequence 21, Application US/09030619B  
Patent No. US20020035061A1  
GENERAL INFORMATION:  
APPLICANT: Krieger, Timothy J.  
APPLICANT: Taylor, Robert  
APPLICANT: Erile, Douglas  
APPLICANT: Fraser, Janet R.  
APPLICANT: West, Michael H.P.  
APPLICANT: McNicol, Patricia J.  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATING  
TITLE OF INVENTION: INJECTIONS USING CATIONIC PEPTIDES ALONE OR IN COMBINATION  
FILE REFERENCE: 660081.406  
CURRENT APPLICATION NUMBER: US/09/030,619B  
CURRENT FILING DATE: 1998-02-25  
NUMBER OF SEQ ID NOS: 232  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 211  
LENGTH: 23  
TYPE: PRT  
ORGANISM: Xenopus laevis  
US-09-030-619-211

Query Match 58.3%; Score 63; DB 10; Length 23;  
Best Local Similarity 100.0%; Pred. No. 0.0018;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 9 GIGKFLHSAKFF 20  
DB 1 GIGKFLHSAKFF 12

RESULT 6  
US-09-904-753-13  
Sequence 13, Application US/09904753  
Publication No. US20030092612A1  
GENERAL INFORMATION:  
APPLICANT: Lynos, Robert T  
TITLE OF INVENTION: Use of Antimicrobial Peptides as Preservatives in  
TITLE OF INVENTION: Ophthalmic Preparations, Including Solutions,  
FILE REFERENCE: 2973 ver 2  
CURRENT APPLICATION NUMBER: US/09/904,753  
CURRENT FILING DATE: 2001-07-13  
PRIOR APPLICATION NUMBER: WO 96/25183  
PRIOR FILING DATE: 1996-08-22  
NUMBER OF SEQ ID NOS: 14  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 13  
LENGTH: 26  
TYPE: PRT  
ORGANISM: Artificial Sequence

FEATURE:  
NAME/KEY: PEPTIDE  
LOCATION: (26)  
OTHER INFORMATION: Xaa at position 26 is Ser-amide  
OTHER INFORMATION: Description of Artificial Sequence: Hybrid  
OTHER INFORMATION: antimicrobial peptide  
US-09-904-753-13

Query Match 58.3%; Score 63; DB 9; Length 26;  
Best Local Similarity 85.7%; Pred. No. 0.002;  
Matches 12; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 KWKLFKKIGIGKFL 14  
DB 1 KWKLFKKIGIGAVL 14

RESULT 7  
US-09-030-619-120  
Sequence 120, Application US/09030619B  
Patent No. US20020035061A1  
GENERAL INFORMATION:  
APPLICANT: Krieger, Timothy J.  
APPLICANT: Taylor, Robert  
APPLICANT: Erfile, Douglas  
APPLICANT: Fraser, Janet R.  
APPLICANT: West, Michael H.P.  
APPLICANT: McNicol, Patricia J.  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATING  
TITLE OF INVENTION: INFECTIONS USING CATIONIC PEPTIDES ALONE OR IN COMBINATION  
FILE REFERENCE: 660081.406  
CURRENT APPLICATION NUMBER: US/09/030,619B  
CURRENT FILING DATE: 1998-02-25  
NUMBER OF SEQ ID NOS: 232  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 120  
LENGTH: 26  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Cationic Peptide Analogue  
US-09-030-619-120

Query Match 58.3%; Score 63; DB 10; Length 26;  
Best Local Similarity 85.7%; Pred. No. 0.002;  
Matches 12; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 KWKLFKKIGIGKFL 14  
DB 1 KWKLFKKIGIGAVL 14

RESULT 8  
US-09-908-139-1  
Sequence 1, Application US/09908139  
Publication No. US20030096949A1  
GENERAL INFORMATION:  
APPLICANT: Hancock, Robert E. W.  
APPLICANT: Gough, Monisha A.  
APPLICANT: Patrzykat, Aleksander  
APPLICANT: Woods, Donald  
APPLICANT: Jia, Xiaoyan  
TITLE OF INVENTION: ANTI-ENDOTOXIC, ANTIMICROBIAL CATIONIC  
TITLE OF INVENTION: PEPTIDES AND METHODS OF USE THEREFOR  
FILE REFERENCE: 07422/016001  
CURRENT APPLICATION NUMBER: US/09/908,139  
CURRENT FILING DATE: 2001-07-17  
PRIOR APPLICATION NUMBER: 09/143,124  
PRIOR FILING DATE: 1998-08-28  
NUMBER OF SEQ ID NOS: 25  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 1

LENGTH: 28  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: antimicrobial cationic peptide  
US-09-908-139-1

Query Match 58.3%; Score 63; DB 9; Length 28;  
Best Local Similarity 85.7%; Pred. No. 0.0022;  
Matches 12; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 KWKLFKKIGIGKFL 14  
DB 1 KWKLFKKIGIGAVL 14

RESULT 9  
US-09-909-652-7  
Sequence 7, Application US/09909652  
Patent No. US20020025537A1  
GENERAL INFORMATION:  
APPLICANT: Kairos Scientific, Inc.  
APPLICANT: Bylina, Edward J.  
APPLICANT: Coleman, William J.  
APPLICANT: Youvan, Douglas C.  
TITLE OF INVENTION: HIGH-THROUGHPUT METHODS FOR GENERATING  
TITLE OF INVENTION: AND SCREENING COMPOUNDS THAT AFFECT CELL VIABILITY  
FILE REFERENCE: 22346-7001  
CURRENT APPLICATION NUMBER: US/09/909,652  
CURRENT FILING DATE: 2001-10-15  
PRIOR APPLICATION NUMBER: US 60/219,179  
PRIOR FILING DATE: 2000-07-19  
NUMBER OF SEQ ID NOS: 7  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 7  
LENGTH: 28  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: CEMA analogue of cecropin-melittin hybrid  
US-09-909-652-7

Query Match 58.3%; Score 63; DB 10; Length 28;  
Best Local Similarity 85.7%; Pred. No. 0.0022;  
Matches 12; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 KWKLFKKIGIGKFL 14  
DB 1 KWKLFKKIGIGAVL 14

RESULT 10  
US-09-030-619-121  
Sequence 121, Application US/09030619B  
Patent No. US20020035061A1  
GENERAL INFORMATION:  
APPLICANT: Krieger, Timothy J.  
APPLICANT: Taylor, Robert  
APPLICANT: Erfile, Douglas  
APPLICANT: Fraser, Janet R.  
APPLICANT: West, Michael H.P.  
APPLICANT: McNicol, Patricia J.  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATING  
TITLE OF INVENTION: INFECTIONS USING CATIONIC PEPTIDES ALONE OR IN COMBINATION  
FILE REFERENCE: 660081.406  
CURRENT APPLICATION NUMBER: US/09/030,619B  
CURRENT FILING DATE: 1998-02-25  
NUMBER OF SEQ ID NOS: 232  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 121  
LENGTH: 28  
TYPE: PRT

ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Cationic Peptide Analogue  
US-09-030-619-121

Query Match 58.3%; Score 63; DB 10; Length 28;  
Best Local Similarity 85.7%; Pred. No. 0.0022;  
Matches 12; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 KMKLFKIGIGKFL 14  
DB 1 KMKLFKIGIGAVL 14

RESULT 11  
US-09-917-340-4  
Sequence 4, Application US/09917340  
Patent No. US20020090369A1  
GENERAL INFORMATION:  
APPLICANT: Murphy, Christopher J.  
APPLICANT: McAnulty, Jonathan F.  
APPLICANT: Reid, Ted W.  
TITLE OF INVENTION: Transplant Media  
FILE REFERENCE: TPLANT-06468  
CURRENT APPLICATION NUMBER: US/09/917,340  
CURRENT FILING DATE: 2001-07-29  
PRIOR APPLICATION NUMBER: 60/221,632  
PRIOR FILING DATE: 2000-07-28  
PRIOR APPLICATION NUMBER: 60/249,602  
PRIOR FILING DATE: 2000-11-17  
PRIOR APPLICATION NUMBER: 60/290,932  
PRIOR FILING DATE: 2001-05-15  
NUMBER OF SEQ ID NOS: 96  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 4  
LENGTH: 303  
TYPE: PRT  
ORGANISM: Xenopus laevis  
US-09-917-340-4

Query Match 58.3%; Score 63; DB 10; Length 303;  
Best Local Similarity 100.0%; Pred. No. 0.024;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 GIGKFLHSAKKF 20  
DB 83 GIGKFLHSAKKF 94

RESULT 12  
US-09-820-053A-146  
Sequence 146, Application US/09820053A  
Publication No. US20030083243A1  
GENERAL INFORMATION:  
APPLICANT: Owen, Donald R.  
TITLE OF INVENTION: SHORT BIOACTIVE PEPTIDES  
FILE REFERENCE: HELIX027  
CURRENT APPLICATION NUMBER: US/09/820,053A  
CURRENT FILING DATE: 2001-03-28  
NUMBER OF SEQ ID NOS: 165  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 146  
LENGTH: 23  
TYPE: PRT  
ORGANISM: ARTIFICIAL SEQUENCE  
FEATURE:  
OTHER INFORMATION: SYNTHETIC SEQUENCE  
NAME/KEY: MOD\_RES  
LOCATION: (23)  
OTHER INFORMATION: AMIDATION  
US-09-820-053A-146

Query Match 55.6%; Score 60; DB 9; Length 23;

Best Local Similarity 91.7%; Pred. No. 0.0054;  
Matches 11; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 9 GIGKFLHSAKKF 20  
DB 1 GIGKFLHSAKKF 12

RESULT 13  
US-09-908-139-9  
Sequence 9, Application US/09908139  
Publication No. US20030096949A1  
GENERAL INFORMATION:  
APPLICANT: Hancock, Robert E. W.  
APPLICANT: Gough, Monisha A.  
APPLICANT: Patrzykat, Aleksander  
APPLICANT: Woods, Donald  
APPLICANT: Jia, Xiaoyan  
TITLE OF INVENTION: ANTI-ENDOTOXIC, ANTIMICROBIAL CATIONIC  
PEPTIDES AND METHODS OF USE THEREFOR  
FILE REFERENCE: 07422/016001  
CURRENT APPLICATION NUMBER: US/09/908,139  
CURRENT FILING DATE: 2001-07-17  
PRIOR APPLICATION NUMBER: 09/143,124  
PRIOR FILING DATE: 1998-08-28  
NUMBER OF SEQ ID NOS: 25  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 9  
LENGTH: 33  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: antimicrobial cationic peptide  
US-09-908-139-9

Query Match 54.6%; Score 59; DB 9; Length 33;  
Best Local Similarity 70.6%; Pred. No. 0.011;  
Matches 12; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 WKLFKIGIGKFLHSAK 18  
DB 3 WKLFKIGIGAVLKVK 19

RESULT 14  
US-09-904-753-1  
Sequence 1, Application US/09904753  
Publication No. US20030092612A1  
GENERAL INFORMATION:  
APPLICANT: Lynos, Robert T.  
TITLE OF INVENTION: Use of Antimicrobial Peptides as Preservatives in  
Ophthalmic Preparations, Including Solutions,  
Emulsions, and Suspensions  
FILE REFERENCE: 2973 ver 2  
CURRENT APPLICATION NUMBER: US/09/904,753  
CURRENT FILING DATE: 2001-07-13  
PRIOR APPLICATION NUMBER: WO 96/25183  
PRIOR FILING DATE: 1996-08-22  
NUMBER OF SEQ ID NOS: 14  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 1  
LENGTH: 23  
TYPE: PRT  
ORGANISM: Xenopus laevis  
PUBLICATION INFORMATION:  
AUTHORS: Lee et al.  
TITLE: High-Level Expression of Antimicrobial Peptide Mediated  
TITLE: by a Fusion Partner Reinforcing Formation of Inclusion  
TITLE: Bodies  
JOURNAL: Biochem. Biophys. Res. Commun.  
VOLUME: 277  
PAGES: 575-580  
DATE: Sept 21, 2000



US-09-904-753-1

Query Match 51.9%; Score 56; DB 9; Length 23;  
 Best Local Similarity 91.7%; Pred. No. 0.024;  
 Matches 11; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 9 GIGKFLHSAGKF 20  
 |||||  
 Db 1 GIGKFLHSAGKF 12

RESULT 15

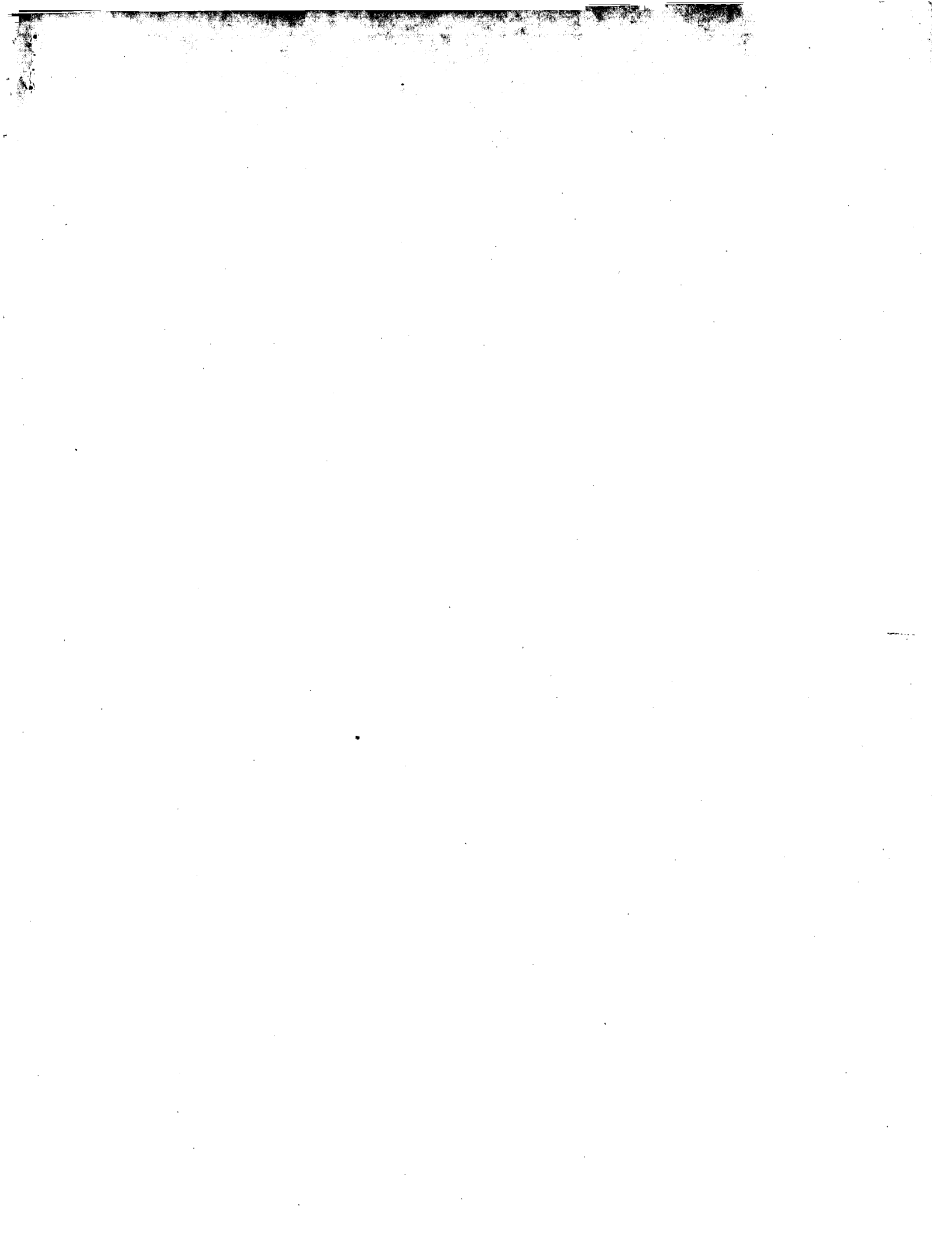
US-09-030-619-210  
 ; Sequence 210, Application US/09030619B  
 ; Patent No. US20020035061A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Krieger, Timothy J.  
 ; APPLICANT: Taylor, Robert  
 ; APPLICANT: Erile, Douglas  
 ; APPLICANT: Fraser, Janet R.  
 ; APPLICANT: West, Michael H.P.  
 ; APPLICANT: McNicol, Patricia J.  
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATING  
 ; TITLE OF INVENTION: INFECTIONS USING CATIONIC PEPTIDES ALONE OR IN COMBINATION  
 ; FILE REFERENCE: 660081.406  
 ; CURRENT APPLICATION NUMBER: US/09/030, 619B  
 ; NUMBER OF SEQ ID NOS: 232  
 ; SOFTWARE: FastSeq for Windows Version 3.0  
 ; SEQ ID NO 210  
 ; LENGTH: 23  
 ; TYPE: PRT  
 ; ORGANISM: Xenopus laevis  
 US-09-030-619-210

Query Match 51.9%; Score 56; DB 10; Length 23;  
 Best Local Similarity 91.7%; Pred. No. 0.024;  
 Matches 11; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 9 GIGKFLHSAGKF 20  
 |||||  
 Db 1 GIGKFLHSAGKF 12

Search completed: June 6, 2003, 13:26:02  
 Job time : 15.5 secs

3 Burkhead  
 Tylenol



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 6, 2003, 13:20:15 ; Search time 30.5 Seconds

(without alignments)  
87.377 Million cell updates/sec

Title: US-10-081-418-2

Perfect score: 105

Sequence: 1 KMKKLLKKPPPLKLLKLL 20

Scoring table: BLOSUM62

Searched: 908470 segs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

```

1: /SID52/gcgdata/geneseq/emb1/AA1980.DAT:*
2: /SID52/gcgdata/geneseq/emb1/AA1981.DAT:*
3: /SID52/gcgdata/geneseq/emb1/AA1982.DAT:*
4: /SID52/gcgdata/geneseq/emb1/AA1983.DAT:*
5: /SID52/gcgdata/geneseq/emb1/AA1984.DAT:*
6: /SID52/gcgdata/geneseq/emb1/AA1985.DAT:*
7: /SID52/gcgdata/geneseq/emb1/AA1986.DAT:*
8: /SID52/gcgdata/geneseq/emb1/AA1987.DAT:*
9: /SID52/gcgdata/geneseq/emb1/AA1988.DAT:*
10: /SID52/gcgdata/geneseq/emb1/AA1989.DAT:*
11: /SID52/gcgdata/geneseq/emb1/AA1990.DAT:*
12: /SID52/gcgdata/geneseq/emb1/AA1991.DAT:*
13: /SID52/gcgdata/geneseq/emb1/AA1992.DAT:*
14: /SID52/gcgdata/geneseq/emb1/AA1993.DAT:*
15: /SID52/gcgdata/geneseq/emb1/AA1994.DAT:*
16: /SID52/gcgdata/geneseq/emb1/AA1995.DAT:*
17: /SID52/gcgdata/geneseq/emb1/AA1996.DAT:*
18: /SID52/gcgdata/geneseq/emb1/AA1997.DAT:*
19: /SID52/gcgdata/geneseq/emb1/AA1998.DAT:*
20: /SID52/gcgdata/geneseq/emb1/AA1999.DAT:*
21: /SID52/gcgdata/geneseq/emb1/AA2000.DAT:*
22: /SID52/gcgdata/geneseq/emb1/AA2001.DAT:*
23: /SID52/gcgdata/geneseq/emb1/AA2002.DAT:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	67	63.8	18	21	AAH17123
2	67	63.8	18	23	ABR73201
3	64	61.0	30	13	AAH1399
4	64	61.0	30	13	AAH22883
5	64	61.0	36	13	AAH21400
6	64	61.0	36	13	AAH22884
7	61	58.1	26	13	AAH21398
8	61	58.1	26	13	AAH22882
9	56.5	53.8	15	15	AAH5957
10	56.5	53.8	16	23	ABH74249

11	56.5	53.8	40	16	AAH84924	Alpha-helix-formin
12	55	52.4	16	19	AAH66517	Amphiphilic peptid
13	55	52.4	17	13	AAH21385	Sequence of amphip
14	55	52.4	17	13	AAH22869	Amphiphilic peptid
15	55	52.4	18	13	AAH21360	Sequence of amphip
16	55	52.4	18	13	AAH21361	Sequence of amphip
17	55	52.4	18	13	AAH22844	Amphiphilic peptid
18	55	52.4	18	13	AAH22844	Amphiphilic peptid
19	55	52.4	18	14	AAH45124	Amphiphilic peptid
20	55	52.4	18	14	AAH31965	Amphiphilic peptid
21	55	52.4	18	14	AAH31155	C-terminal subseq.
22	55	52.4	18	14	AAH35373	Amphiphilic peptid
23	55	52.4	18	15	AAH55979	Ion channel formin
24	55	52.4	18	15	AAH50440	Amphiphilic peptid
25	55	52.4	18	15	AAH50555	Amphiphilic peptid
26	55	52.4	18	15	AAH56940	Amphiphilic peptid
27	55	52.4	18	15	AAH59057	Peptide which neut
28	55	52.4	18	16	AAH90251	Cancer treating, a
29	55	52.4	18	19	AAH66488	Amphiphilic peptid
30	55	52.4	18	20	AAH10742	Amphiphilic peptid
31	54	51.4	18	23	AAH23455	Peptide used to ma
32	54	51.4	18	13	AAH21351	Biologically activ
33	53	50.5	17	19	AAH22835	Sequence of amphip
34	53	50.5	17	19	AAH71679	Amphiphilic peptid
35	53	50.5	17	22	AAH50847	Model calmodulin b
36	52.5	50.0	114	21	AAH23000	Synthetic calmodul
37	52.5	50.0	114	21	AAH47567	Arabidopsis thalia
38	52.5	50.0	115	21	AAH22999	Arabidopsis thalia
39	52.5	50.0	115	21	AAH47566	Arabidopsis thalia
40	52.5	50.0	124	21	AAH47540	Arabidopsis thalia
41	52.5	50.0	124	21	AAH47539	Arabidopsis thalia
42	52.5	50.0	173	21	AAH22998	Arabidopsis thalia
43	52.5	50.0	173	21	AAH47565	Arabidopsis thalia
44	52.5	50.0	183	21	AAH47538	Arabidopsis thalia
45	52	49.5	14	13	AAH21402	Sequence of amphip
			14	14	AAH45136	Amphiphilic peptid

#### ALIGNMENTS

```

RESULT 1
ID      AAB17123 standard; Peptide: 18 AA.
AC      AAB17123;
XX      31-OCT-2000 (first entry)
DE      Calmodulin antagonist peptide sequence SEQ ID NO:179.
KW      Modified peptide; therapeutic agent; fusion; Fc domain; cancer.
KW      autoimmune disease; cytostatic; antitumoral; thrombolytic; VEGF;
KW      immunosuppressive; EPO; TPO; CTLA4; miltetic; IL-1; TNF; antagonist;
KW      MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
KW      cytotoxic T cell lymphocyte antigen 4; tumor necrosis factor;
KW      vascular endothelial growth factor; matrix metalloproteinase;
KW      asthma; thrombosis; pharmaceutical.
XX      Synthetic.
XX      OS
XX      PN      WO200024782-A2.
XX      PD      04-MAY-2000.
XX      PF      25-OCT-1999; 99WO-US25044.
XX      PR      23-OCT-1998; 98US-0105371.
XX      PR      22-OCT-1999; 99US-0428082.
XX      PA      (AMGE-) AMGEN INC.
XX      Feige U, Liu C, Cheetham J, Boone TC;
XX

```

DR WPI; 2000-350702/30.  
XX  
PT Novel composition of matter comprising an Fc domain and  
PT pharmacologically active peptides, useful for treating cancer and  
PT autoimmune diseases -  
XX  
PS Claim 39; Page 255; 608pp; English.  
XX  
CC The present invention describes composition of matter (I) comprising an  
CC Fc domain, pharmacologically active peptides, and linkers. Where (I) is:  
CC (X1)-a-F1-(X2)-b, where: F1 - an Fc domain; X1 and X2 - are each  
CC independently selected from -(L1)-C-P1, -(L1)-C-P1-(L2)-d-P2,  
CC -(L1)-C-P1-(L2)-e-P3, or -(L1)-C-P1-(L2)-d-P2-(L3)-e-P3-(L4)-f-P4  
CC where P1, P2, P3, and P4 - are each independently sequences of  
CC pharmacologically active peptides; L1, L2, L3, and L4 - are each  
CC independently linkers; and a, b, c, d, e, and f - are each independently  
CC 0 or 1, provided that at least 1 of a and b is 1. The composition can  
CC have cytostatic, antineoplastic, thrombolytic and immunosuppressive  
CC activities. DNAs, vectors and host cells from the present invention can  
CC be used for producing pharmaceutical compositions. The compositions are  
CC useful for treating cancer, asthma, thrombosis, or autoimmune diseases.  
CC The use of an Fc domain (rather than a Fab domain) can provide a longer  
CC half-life or incorporate functions such as Fc receptor binding, protein  
CC A binding, complement fixation, and possibly placental transfer. AAB69443  
CC to AAB69556 and AAB16955 to AAB18003 represent nucleotide and amino acid  
CC sequences used in the exemplification of the present invention.  
XX  
SQ Sequence 18 AA;  
XX  
Query Match 63.8%; Score 67; DB 21; Length 18;  
Best Local Similarity 80.0%; Pred. No. 0.03;  
Matches 16; Conservative 0; Mismatches 0; Indels 4; Gaps 1;  
OY 1 KWKRLKKPPPLKLLKLL 20  
1 ||||| |||||  
Db 2 KWKRLK---LKKLLKLL 17  
XX  
RESULT 2  
AAB73201  
ID AAB73201 standard; Peptide: 18 AA.  
XX  
AC AAB73201;  
XX  
DF 05-APR-2002 (first entry)  
XX  
DE Calmodulin antagonist peptide SEQ ID NO:179.  
XX  
KW Modified peptide; mimetic; Fc domain; fusion; immunoglobulin G; IgG;  
KW EPO; erythropoietin; TPO; tumour necrosis factor alpha inhibitor;  
KW TNF-alpha inhibitor; interleukin 1 antagonist; IL-1 antagonist; TNP;  
KW TPO mimetic peptide; EPO mimetic peptide; EMP; VEGF antagonist;  
KW MMP inhibitor; antiinflammatory; antitumour; immunosuppressive;  
KW cytostatic; antineoplastic; antiarthritic; antidiabetic; ophthalmological;  
KW antianemic; anorectic; antifertility; haemostatic; dermatological;  
KW neuroprotective; inflammatory disease; autoimmune disease; tumour growth;  
KW cancer; rheumatoid arthritis; diabetic retinopathy; infertility; obesity;  
KW sleep disorder; neurological degenerative disease; anaemia;  
KW thrombocytopaenia; metastatic tumour; systemic lupus erythematosus;  
KW Fanconi's syndrome.  
XX  
OS Homo sapiens.  
OS Synthetic.  
XX  
PN WO200183525-A2.  
XX  
PD 08-NOV-2001.  
XX  
PF 02-MAY-2001; 2001WO-US14310.  
XX  
PR 03-MAY-2000; 2000US-0563286.  
XX  
PA (AMGE-) AMGEN INC.

XX  
PI Feige U, Liu C, Cheetham JC, Boone TC, Gudas JM;  
XX  
DR WPI; 2002-130313/17.  
XX  
PT Novel vehicle-peptide molecule or its multimers useful for treating  
PT inflammatory and autoimmune diseases, cancer, rheumatoid arthritis,  
PT diabetic retinopathy, obesity, sleep disorders and infertility -  
XX  
PS Claim 39; Page 53; 176pp; English.  
XX  
CC The present invention describes a vehicle-peptide molecule (I) or its  
CC multimers. (I) can have antiinflammatory, antitumour, immunosuppressive,  
CC cytostatic, antineoplastic, antiarthritic, antidiabetic, ophthalmological,  
CC antianemic, anorectic, antifertility, haemostatic, dermatological and  
CC neuroprotective activities. (I) can be used as a therapeutic or  
CC prophylactic agent as well as for screening purposes. (I) is useful for  
CC diagnosing diseases characterised by dysfunction or abnormal proteins of  
CC protein of interest, for identifying normal or abnormal proteins of  
CC interest, as a part of diagnostic kit to detect the presence of their  
CC proteins of interest in a biological sample. Additionally, (I) is useful  
CC for treating inflammatory and autoimmune diseases, tumour growth, cancer,  
CC rheumatoid arthritis, diabetic retinopathy, obesity, sleep disorders,  
CC infertility, and neurological degenerative diseases. (I), comprising  
CC EPO-mimetic compounds are useful for treating disorders characterised by  
CC low red blood cell levels such as anaemia. The TPO-mimetic comprising  
CC compounds are useful for treating conditions that involve an existing  
CC megakaryocyte/platelet deficiency or an expected megakaryocyte/platelet  
CC deficiency, such as thrombocytopaenia, aplastic anaemia, metastatic  
CC tumour which result in thrombocytopaenia, systemic lupus erythematosus,  
CC and Fanconi's syndrome. AAB72403 to AAB73426 and AAB135695 to AAB135777  
CC represent amino acid and nucleic acid sequences used in the  
CC exemplification of the present invention.  
XX  
SQ Sequence 18 AA;  
XX  
Query Match 63.8%; Score 67; DB 23; Length 18;  
Best Local Similarity 80.0%; Pred. No. 0.03;  
Matches 16; Conservative 0; Mismatches 0; Indels 4; Gaps 1;  
OY 1 KWKRLKKPPPLKLLKLL 20  
1 ||||| |||||  
Db 2 KWKRLK---LKKLLKLL 17  
XX  
RESULT 3  
AAR21399  
ID AAR21399 standard; peptide: 30 AA.  
XX  
AC AAR21399;  
XX  
DT 16-MAY-1992 (first entry)  
XX  
DE Sequence of amphiphilic peptide SEQ ID NO. 62 with  
DE acetylated N-terminus, may be a C-terminal amide.  
XX  
KW Amphiphilic peptide; anti-microbial; anti-viral; anti-tumour;  
KW spermicide; wound healing; steriliant.  
XX  
PN WO9201462-A.  
XX  
PD 06-FEB-1992.  
XX  
PF 17-JUL-1991; 91WO-US05047.  
XX  
PR 19-JUL-1990; 90US-0554422.  
XX  
PA (SCRI-) SCRIPPS RES INST.  
XX  
PI Houghten RA, Blondelle S;  
XX  
DR WPI; 1992-064700/08.

PT Method for inhibiting target cell and virus growth - comprises  
PT administering amphiphilic peptide compns, useful for treating  
PT viral and phytopathogenic infections, tumours and burns  
XX  
PS Disclosure; Page 60; 72pp; English.  
XX  
CC The peptides of the invention are effective pharmaceuticals having  
CC anti-microbial, anti-viral and anti-tumour activity. They are also  
CC useful for inhibiting, preventing or destroying the motility of  
CC sperm and hence have application in a spermicide preparation. They  
CC also have anti-parasitic activity and are useful in wound healing.  
CC as preservatives and sterilants and to inhibit growth of  
CC phytopathogenic microorganisms. AAR20969 and AAR20970 were published in  
CC Haighten and Ostresh, Bio Chromatography, Vol 2, Issue 2, page 80-83,  
CC 1987.  
XX  
SQ Sequence 30 AA;  
XX  
Query Match 61.0%; Score 64; DB 13; Length 30;  
Best Local Similarity 80.0%; Pred. No. 0.12;  
Matches 16; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
QY 1 KMKKLLKKPPPLKKLKKL 20  
Db 4 KKKKLLKKLKKLKKLKKL 23  
XX  
RESULT 4  
AAR22883  
ID AAR22883 standard; Peptide; 30 AA.  
XX  
AC AAR22883;  
XX  
DT 22-AUG-1992 (first entry)  
XX  
DE Amphiphilic peptide to inhibit growth of a target cell.  
XX  
KM Antimicrobial; antifungal; antitumour; pharmaceuticals; antibiotics;  
KM antiparasitic; spermicides; burns; wound healing.  
XX  
OS Synthetic.  
XX  
FH Key Location/Qualifiers  
FT Modified-site 1  
FT Modified-site /note="acylated"  
FT Modified-site 30  
FT Modified-site /note="amidated"  
XX  
PN CA2047317-A.  
XX  
PD 20-JAN-1992.  
XX  
PE 18-JUL-1991; 91CA-2024317.  
XX  
PR 08-JUL-1991; 91US-0725331.  
PR 19-JUL-1990; 90US-0554442.  
XX  
PA (SCRI-) SCRIPPS RES INST.  
XX  
PI Houghten RA, Blondelle S;  
PI WPI; 1992-114943/15.  
XX  
DR  
XX  
XX Amphiphilic peptide(s) and analogues - for use in e.g.  
PT antimicrobial, antifungal or antitumour compositions, having  
PT increased biological activity  
XX  
PS Claim 30; Page 60; 71pp; English.  
XX  
CC The amphiphilic peptide (SEQ ID NO 62) was prepd. by standard  
CC solid phase methods. The peptide is an analogue of the peptide  
CC having SEQ ID NO 3 (AAR22824), with an N-terminal extension of 12  
CC amino acids. Substitution and deletion analogues of this peptide

CC have increased biological activity and are effective as pharma-  
CC ceuticals e.g. antibiotics for bacterial, fungal or viral  
CC infections, or in spermicides or antitumour or antiparasitic  
CC agents. Additionally the peptides can be used in wound healing  
CC compns. or for treating burns or other skin or eye infections.  
CC See also AAR2822-89.  
XX  
SQ Sequence 30 AA;  
XX  
Query Match 61.0%; Score 64; DB 13; Length 30;  
Best Local Similarity 80.0%; Pred. No. 0.12;  
Matches 16; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
QY 1 KMKKLLKKPPPLKKLKKL 20  
Db 4 KKKKLLKKLKKLKKLKKL 23  
XX  
RESULT 5  
AAR21400  
ID AAR21400 standard; peptide; 36 AA.  
XX  
AC AAR21400;  
XX  
DT 16-MAY-1992 (first entry)  
XX  
DE Sequence of amphiphilic peptide SEQ ID NO. 63 with  
DE acetylated N-terminus, may be a C-terminal amide.  
XX  
KM Amphiphilic peptide; anti-microbial; anti-viral; anti-tumour;  
KM spermicide; wound healing; sterilant.  
XX  
PN WO9201462-A.  
XX  
PD 06-FEB-1992.  
XX  
PE 17-JUL-1991; 91WO-US05047.  
XX  
PR 19-JUL-1990; 90US-0554422.  
XX  
PA (SCRI-) SCRIPPS RES INST.  
XX  
PI Houghten RA, Blondelle S;  
PI WPI; 1992-064700/08.  
XX  
DR  
XX  
PT Method for inhibiting target cell and virus growth - comprises  
PT administering amphiphilic peptide compns, useful for treating  
PT viral and phytopathogenic infections, tumours and burns  
XX  
PS Disclosure; Page 61; 72pp; English.  
XX  
CC The peptides of the invention are effective pharmaceuticals having  
CC anti-microbial, anti-viral and anti-tumour activity. They are also  
CC useful for inhibiting, preventing or destroying the motility of  
CC sperm and hence have application in a spermicide preparation. They  
CC also have anti-parasitic activity and are useful in wound healing,  
CC as preservatives and sterilants and to inhibit growth of  
CC phytopathogenic microorganisms. AAR20969 and AAR20970 were published in  
CC Haighten and Ostresh, Bio Chromatography, Vol 2, Issue 2, page 80-83,  
CC 1987.  
XX  
SQ Sequence 36 AA;  
XX  
Query Match 61.0%; Score 64; DB 13; Length 36;  
Best Local Similarity 80.0%; Pred. No. 0.14;  
Matches 16; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
QY 1 KMKKLLKKPPPLKKLKKL 20  
Db 10 KKKKLLKKLKKLKKLKKL 29

```

RESULT 6
AAR22884 ID AAR22884 standard; Peptide; 36 AA.
XX
AC AAR22884:
XX
DT 22-AUG-1992 (first entry)
XX
DE Amphiphilic peptide to inhibit growth of a target cell.
XX
KW Antimicrobial; antifungal; antitumour; pharmaceuticals; antibiotics;
XX antiparasitic; spermicides; burns; wound healing.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Modified-site 1 /note="acylated"
FT Modified-site 36 /note="amidated"
FT
XX CA2047317-A.
XX
PD 20-JAN-1992.
XX
PF 18-JUL-1991; 91CA-2024317.
XX
PR 08-JUL-1991; 91US-0725331.
PR 19-JUL-1990; 90US-0554442.
XX
PA (SCRI-) SCRIPPS RES INST.
XX
PI Houghten RA, Blondelle S;
XX
DR WPI: 1992-114943/15.
XX
PT Amphiphilic peptide(s) and analogues - for use in e.g.
PT antimicrobial, antifungal or antitumour compositions, having
PT increased biological activity
XX
XX Claim 30; Page 61; 71pp; English.
XX
CC The amphiphilic peptide (SEQ ID NO 63) was prepd. by standard
CC solid phase methods. The peptide is an analogue of the peptide
CC having SEQ ID NO 3 (AAR22824), with an N-terminal extension of 18
CC amino acids. Substitution and deletion analogues of this peptide
CC have increased biological activity and are effective as pharma-
CC ceuticals e.g. antibiotics for bacterial, fungal or viral
CC infections, or in spermicides or antitumour or antiparasitic
CC agents. Additionally the peptides can be used in wound healing
CC compns. or for treating burns or other skin or eye infections.
CC See also AAR2282-89.
XX
SQ Sequence 36 AA:
Query Match 61.0%; Score 64; DB 13; Length 36;
Best Local Similarity 80.0%; Pred. No. 0.14;
Matches 16; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 1 KKKLKKPPPLKKLKKL 20
DB 10 KKKLKKLKKLKKLKKL 29
RESULT 7
AAR21398 ID AAR21398 standard; peptide; 26 AA.
XX
AC AAR21398:
XX
DT 16-MAY-1992 (first entry)
XX
DE Sequence of amphiphilic peptide SEQ ID No. 61 with

```

```

DE acetylated N-terminus, may be a C-terminal amide.
XX
XX Amphiphilic peptide; anti-microbial; anti-viral; anti-tumour;
KW spermicide; wound healing; sterlant.
XX
XX WO9201462-A.
XX
PD 06-FEB-1992.
XX
PF 17-JUL-1991; 91WO-US05047.
XX
PR 19-JUL-1990; 90US-0554422.
XX
PA (SCRI-) SCRIPPS RES INST.
XX
PI Houghten RA, Blondelle S;
XX
DR WPI: 1992-064700/08.
XX
PT Method for inhibiting target cell and virus growth - comprises
PT administering amphiphilic peptide compns, useful for treating
PT viral and phytopathogenic infections, tumours and burns
XX
PS Disclosure; Page 60; 72pp; English.
XX
CC The peptides of the invention are effective pharmaceuticals having
CC anti-microbial, anti-viral and anti-tumour activity. They are also
CC useful for inhibiting, preventing or destroying the motility of
CC sperm and hence have application in a spermicide preparation. They
CC also have anti-parasitic activity and are useful in wound healing,
CC as preservatives and sterilants and to inhibit growth of
CC phytopathogenic microorganisms. AAR20969 and AAR20970 were published in
CC Houghten and Ostresh, Bio Chromatography, Vol 2, Issue 2, page 80-83,
CC 1987.
XX
SQ Sequence 26 AA:
Query Match 58.1%; Score 61; DB 13; Length 26;
Best Local Similarity 83.3%; Pred. No. 0.25;
Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 3 KKKLKKPPPLKKLKKL 20
DB 2 KKKLKKLKKLKKLKKL 19
RESULT 8
AAR22882 ID AAR22882 standard; Peptide; 26 AA.
XX
AC AAR22882:
XX
DT 22-AUG-1992 (first entry)
XX
DE Amphiphilic peptide to inhibit growth of a target cell.
XX
KW Antimicrobial; antifungal; antitumour; pharmaceuticals; antibiotics;
KW antiparasitic; spermicides; burns; wound healing.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Modified-site 1 /note="acylated"
FT Modified-site 26 /note="amidated"
FT
XX CA2047317-A.
XX
PD 20-JAN-1992.
XX
PF 18-JUL-1991; 91CA-2024317.
XX

```

08-JUL-1991: 91US-0725331.  
 19-JUL-1990: 90US-0554442.  
 (SCRI-) SCRIPPS RES INST.  
 Houghten RA, Blondelle S;  
 WPI; 1992-114943/15.  
 Amphiphilic peptide(s) and analogues - for use in e.g.  
 antimicrobial, antifungal or antitumour compositions, having  
 increased biological activity  
 Claim 30; Page 60; 71pp; English.  
 The amphiphilic peptide (SEQ ID NO 61) was prepd. by standard  
 solid phase methods. The peptide is an analogue of the peptide  
 having SEQ ID NO 3 (AAR2824), with an N-terminal extension of 8  
 amino acids. Substitution and deletion analogues of this peptide  
 have increased biological activity and are effective as pharma-  
 ceuticals e.g. antibiotics for bacterial, fungal or viral  
 infections, or in spermicides or antitumour or antiparasitic  
 agents. Additionally the peptides can be used in wound healing  
 compns. or for treating burns or other skin or eye infections.  
 See also AAR2822-89.  
 Sequence 26 AA;  
 Query Match 58.1%; Score 61; DB 13; Length 26;  
 Best Local Similarity 83.3%; Pred. No. 0.25;  
 Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 3 KRLKKPPPLKRLKRL 20  
 ||||| |||||  
 Db 2 KRLKKLKLKRLKRL 19  
 ||||| |||||  
 RESULT 9  
 AAR56957  
 ID AAR56957 standard; peptide; 15 AA.  
 AC AAR56957;  
 17-MAR-1995 (first entry)  
 Peptide which neutralises bacterial endotoxin.  
 septic shock; bacterial endotoxin; lipopolysaccharide; LPS;  
 gram negative bacteria; conjugate moiety; septicemia; neutralising;  
 longer activity; polyvinylpyrrolidone; dextran; beta starch;  
 polyvinyl alcohol; ion-channel forming; amphiphilic.  
 Synthetic.  
 WO9413697-A.  
 23-JUN-1994.  
 06-DEC-1993: 93WO-US11841.  
 07-DEC-1992: 92US-0987443.  
 (MAGA-) MAGAININ PHARM INC.  
 Hendi M, Rao M, Williams TJ;  
 WPI; 1994-217804/26.  
 New conjugates of bioactive amphiphilic peptide(s) and conjugate  
 moiety - are useful for treatment of septic shock  
 Disclosure; Page 120; 141pp; English.

Septic shock is often due to the body's reaction to foreign  
 lipopolysaccharide (LPS). The compounds of the invention neutralise  
 bacterial endotoxins without neutralising essential proteins in the  
 plasma of patients, eg. heparins. They also have longer duration of  
 activity than unconjugated peptides. In general peptides such as this  
 are ion-channel forming peptides. The compounds are biologically active  
 peptides linked to a conjugate moiety, eg. carbohydrates, proteins,  
 polyvinylpyrrolidone, polyalkylene glycols and polyvinyl alcohols.  
 The conjugate moiety may be linked at the C- or N-terminal or  
 internally of the peptide. AAR5591-631 and AAR56879-957 are examples  
 of these peptide-conjugate moiety compounds  
 Sequence 15 AA;  
 Query Match 53.8%; Score 56.5; DB 15; Length 15;  
 Best Local Similarity 83.3%; Pred. No. 0.58;  
 Matches 15; Conservative 0; Mismatches 0; Indels 3; Gaps 1;  
 3 KRLKKPPPLKRLKRL 20  
 ||||| |||||  
 Db 1 KRLKK--LKLKRL 15  
 |||||  
 RESULT 10  
 ABB74249  
 ID ABB74249 standard; Peptide; 16 AA.  
 AC ABB74249;  
 18-APR-2002 (first entry)  
 Amphiphilic fusogenic peptide SEQ ID NO:15.  
 Fusogenic; nuclear localisation signal; NLS; encapsulation; lipogene;  
 liposome; micelle; karyophilic; cytosolic; antitumour; solid tumour;  
 peptide-lipid-poly nucleotide complex; neoplastic disease; gene therapy;  
 breast carcinoma; prostate carcinoma.  
 Synthetic.  
 WO200193836-A2.  
 13-DEC-2001.  
 08-JUN-2001: 2001WO-US18657.  
 09-JUN-2000: 2000US-210925P.  
 (BOUL/) BOULIKAS T.  
 Bouliskas T;  
 WPI; 2002-164295/21.  
 Encapsulation of plasmid DNA (lipogenes) and therapeutic agents with  
 nuclear localisation signal/fusogenic peptide conjugates into targeted  
 liposome complexes -  
 Example; Page 38; 107pp; English.  
 The present invention describes a method for producing micelles with  
 entrapped therapeutic agents. The method comprises: (1) combining  
 negatively charged agent with a cationic lipid in a ratio where 30-90 %  
 of the negatively charged atoms are neutralised by positive charges on  
 lipid molecules to form an electrostatic micelle complex in 20-80 %  
 ethanol; and (2) combining the micelle complex of (a) with fusogenic-  
 karyophilic peptide conjugates in a 0.0-0.3 ratio, therefore producing  
 micelles with entrapped therapeutic agents. Also described is a method  
 for delivering a therapeutic agent in vivo, comprising the administration  
 of the micelle. ABB74256 to ABB74858 represent specifically claimed  
 nuclear localisation signal (NLS) peptides for use in the method as the  
 fusogenic-karyophilic peptides. The micelles produced can have cytostatic  
 and antitumour activities. The peptide-lipid-poly nucleotide complexes

CC produced are useful for inhibiting the progression of neoplastic  
 CC diseases. The invention relates to the field of gene therapy and is  
 CC directed toward methods for producing peptide-lipid-poly nucleotide  
 CC complexes suitable for delivery of polynucleotides. The encapsulated  
 CC molecules display therapeutic efficacy in eradicating solid tumours  
 CC including but not limited to breast carcinoma or prostate carcinoma.  
 CC ABB74235 to ABB74255 are used in the exemplification of the present  
 CC invention.

CC XX Sequence 16 AA;

Query Match 53.8%; Score 56.5; DB 23; Length 16;  
 Best Local Similarity 83.3%; Pred. No. 0.61;  
 Matches 15; Conservative 0; Mismatches 0; Indels 3; Gaps 1;

OY 3 KKLKKPPPLKKLKKL 20  
 ||||| |||||  
 Db 2 KKLKK---LKKLKKL 16

RESULT 11  
 AAR84924  
 ID AAR84924 standard; peptide; 40 AA.

AC AAR84924;  
 DT 18-MAR-1996 (first entry)  
 DE Alpha-helix-forming oligopeptide (LKKL)10.

KW Alpha-helix; secondary structure; nucleic acid transfer; cationic;  
 KM DNA binding peptide; gene therapy; encapsulation.

XX OS Synthetic.

XX FT Key Location/Qualifiers  
 FT Region 1..4  
 FT /label= LKKL  
 FT /note= "one of 10 repeat units"

PN FR2715847-A1.

PD 11-AUG-1995.

PE 08-FEB-1994; 94FR-0001381.

PR 08-FEB-1994; 94FR-0001381.

PA (RHON) RHONE POULENC RORER SA.

PI Bazile D, Emile C, Helene C, Spenlehauer G;

DR WPI: 1995-276981/37.

PT Complex of nucleic acid and oligopeptide with sec. structure - and  
 PT transfer vectors contg. them, useful for efficient transfer of  
 PT nucleic acid to cells in gene therapy.

PS Claim 6 and Example 1.1; Page 7; 20pp; French.

XX The present peptide is a specific example of a cationic oligopeptide  
 CC corresp. to the formula (b-1-l-b)n, where b is a hydrophobic amino acid,  
 CC l is a hydrophilic amino acid and n is at least 4. In this case, where  
 CC b is leu, l is lys and n = 10, the oligopeptide forms an alpha-helix  
 CC which forms a stable complex with a nucleic acid. The complex  
 CC is suitable for transferring nucleic acid, esp. in gene therapy.

XX SQ Sequence 40 AA;

Query Match 53.8%; Score 56.5; DB 16; Length 40;  
 Best Local Similarity 83.3%; Pred. No. 1.4;  
 Matches 15; Conservative 0; Mismatches 0; Indels 3; Gaps 1;

OY 3 KKLKKPPPLKKLKKL 20  
 ||||| |||||  
 Db 2 KKLKK---LKKLKKL 16

RESULT 12  
 AAM66517  
 ID AAM66517 standard; peptide; 16 AA.

AC AAM66517;

DT 25-NOV-1998 (first entry)

DE Amphiphilic peptide.

KW magainin; analogue; antimicrobial; antitumour; wound healing;

KW CPE; amphiphilic; XFP peptide.

XX OS Synthetic.

XX FT Key Location/Qualifiers  
 FT Misc-difference 1..16  
 FT /note= "D-form residues"

PN US5792831-A.

PD 11-AUG-1998.

PE 17-NOV-1994; 94US-0343882.

PR 05-OCT-1993; 93US-0133740.

PR 08-FEB-1990; 90US-0476629.

PR 14-MAY-1990; 90US-0522688.

PR 28-APR-1992; 92US-0874685.

PR 17-NOV-1994; 94US-0343882.

PA (MAGA-) MAGAININ PHARM INC.

PI Maloy WL;

DR WPI: 1998-456190/39.

PT Magainin peptide analogues - useful as antimicrobial or antitumour  
 PT agents, etc.

PS Disclosure; Column 18; 25pp; English.

XX The invention relates to analogues of a magainin I or II, D-form  
 CC analogues, deletion analogues or related peptides. It also relates  
 CC to basic polypeptides having at least 16 amino acids, including at least  
 CC 8 hydrophobic amino acids and at least 8 hydrophilic amino acids. The  
 CC peptides may be used as antimicrobial agents, antiviral agents,  
 CC antibiotics, antitumour agents, antiparasitic agents, spermicides,  
 CC preservatives or sterilants, or agents for promoting wound healing. The  
 CC present sequence represents a specific example of a peptide disclosed in  
 CC the specification.

XX SQ Sequence 16 AA;

Query Match 52.4%; Score 55; DB 19; Length 16;  
 Best Local Similarity 75.0%; Pred. No. 0.96;  
 Matches 15; Conservative 0; Mismatches 1; Indels 4; Gaps 1;

OY 1 KKKLKKPPPLKKLKKL 20  
 | ||||| |||||  
 Db 1 KKLKKLKK---LKKLKKL 16

RESULT 13

AAR21385  
 ID AAR21385 standard; peptide; 17 AA.

XX AAR21385;



XX 16-MAY-1992 (first entry)  
 DT XX  
 DE Sequence of amphiphilic peptide SEQ ID No. 48 with  
 DE C-terminal amide and acetylated at N-terminus.  
 XX  
 KW Amphiphilic peptide; anti-microbial; anti-viral; anti-tumour;  
 KW spermicide; wound healing; steriliant.  
 XX  
 PN WO9201462-A.  
 XX  
 PD 06-FEB-1992.  
 XX  
 PF 17-JUL-1991; 91WO-US05047.  
 XX  
 PR 19-JUL-1990; 90US-0554422.  
 XX  
 PA (SCRI-) SCRIPPS RES INST.  
 XX  
 PI Houghten RA, Blondelle S;  
 XX  
 DR WPI: 1992-064700/08.  
 XX  
 PT Method for inhibiting target cell and virus growth - comprises  
 PT administering amphiphilic peptide compns, useful for treating  
 PT viral and phytopathogenic infections, tumours and burns  
 XX  
 PS Claim 4; Page 54; 72pp; English.  
 XX  
 CC The peptides of the invention are effective pharmaceuticals having  
 CC anti-microbial, anti-viral and anti-tumour activity. They are also  
 CC useful for inhibiting, preventing or destroying the motility of  
 CC sperm and hence have application in a spermicide preparation. They  
 CC also have anti-parasitic activity and are useful in wound healing,  
 CC as preservatives and sterilants and to inhibit growth of  
 CC phytopathogenic microorganisms. AAR20969 and AAR20970 were published in  
 CC CC Highten and Ostresh, Bio Chromatography, Vol 2, Issue 2, page 80-83,  
 CC 1987.  
 XX  
 SQ Sequence 17 AA:  
 QY  
 Db 1 KWKLLKKPPPLKKLLKL 20  
 2 KKKLLKK---LKKLLKKL 17

Query Match 52.4%; Score 55; DB 13; Length 17;  
 Best Local Similarity 75.0%; Pred. No. 1;  
 Matches 15; Conservative 0; Mismatches 1; Indels 4; Gaps 1;

RESULT 14  
 AAR22869  
 ID AAR22869 standard; Peptide; 17 AA.  
 XX  
 AC AAR22869;  
 XX  
 DT 22-AUG-1992 (first entry)  
 XX  
 DE Amphiphilic peptide to inhibit growth of a target cell.  
 XX  
 KW Antimicrobial; antifungal; antitumour; pharmaceuticals; antibiotics;  
 KW antiparasitic; spermicides; burns; wound healing.  
 XX  
 OS Synthetic.  
 OS  
 FH Key Location/Qualifiers  
 FH 1  
 FT Modified-site /note= "acylated"  
 FT 17  
 FT Modified-site /note= "amidated"  
 FT  
 PN CA2047317-A.  
 XX

PD 20-JAN-1992.  
 XX  
 PF 18-JUL-1991; 91CA-2024317.  
 XX  
 PR 08-JUL-1991; 91US-0725331.  
 PR 19-JUL-1990; 90US-0554442.  
 XX  
 PA (SCRI-) SCRIPPS RES INST.  
 XX  
 PI Houghten RA, Blondelle S;  
 XX  
 DR WPI: 1992-114943/15.  
 XX  
 PF Amphiphilic peptide(s) and analogues - for use in e.g.  
 PF antimicrobial, antifungal or antitumour compositions, having  
 PF increased biological activity  
 XX  
 PS Claim 9; Page 54; 71pp; English.  
 XX  
 CC The amphiphilic peptide (SEQ ID NO 48) was prepd. by standard  
 CC solid phase methods. The peptide is an analogue of the peptide  
 CC having SEQ ID NO 3 (AAR22824), with Leu at position 4 deleted.  
 CC Substitution and deletion analogues of this peptide have  
 CC increased biological activity and are effective as pharmaceuticals  
 CC e.g. antibiotics for bacterial, fungal or viral infections, or in  
 CC spermicides or antitumour or antiparasitic agents. Additionally  
 CC the peptides can be used in wound healing compns. or for treating  
 CC burns or other skin or eye infections.  
 CC See also AAR22822-89.  
 XX  
 SQ Sequence 17 AA:  
 QY  
 Db 1 KWKLLKKPPPLKKLLKL 20  
 2 KKKLLKK---LKKLLKKL 17

Query Match 52.4%; Score 55; DB 13; Length 17;  
 Best Local Similarity 75.0%; Pred. No. 1;  
 Matches 15; Conservative 0; Mismatches 1; Indels 4; Gaps 1;

RESULT 15  
 AAR21360  
 ID AAR21360 standard; peptide; 18 AA.  
 XX  
 AC AAR21360;  
 XX  
 DT 16-MAY-1992 (first entry)  
 XX  
 DE Sequence of amphiphilic peptide SEQ ID No. 23 which may be a C-  
 DE terminal amide and/or may be acetylated at N-terminus.  
 XX  
 KW Amphiphilic peptide; anti-microbial; anti-viral; anti-tumour;  
 KW spermicide; wound healing; steriliant.  
 XX  
 PN WO9201462-A.  
 XX  
 PD 06-FEB-1992.  
 XX  
 PF 17-JUL-1991; 91WO-US05047.  
 XX  
 PR 19-JUL-1990; 90US-0554422.  
 XX  
 PA (SCRI-) SCRIPPS RES INST.  
 XX  
 PI Houghten RA, Blondelle S;  
 XX  
 DR WPI: 1992-064700/08.  
 XX  
 PT Method for inhibiting target cell and virus growth - comprises  
 PT administering amphiphilic peptide compns, useful for treating  
 PT viral and phytopathogenic infections, tumours and burns  
 XX

PS Claim 4; Page 42; 72pp; English.

XX  
CC The peptides of the invention are effective pharmaceuticals having  
CC anti-microbial, anti-viral and anti-tumour activity. They are also  
CC useful for inhibiting, preventing or destroying the motility of  
CC sperm and hence have application in a spermicide preparation. They  
CC also have anti-parasitic activity and are useful in wound healing,  
CC as preservatives and sterilants and to inhibit growth of  
CC phytopathogenic microorganisms. AAR20969 and AAR20970 were published in  
CC Haighten and Ostresh, Bio Chromatography, Vol 2, Issue 2, page 80-83,  
CC 1987.

XX  
SQ Sequence 18 AA;

Query Match 52.4%; Score 55; DB 13; Length 18;

Best Local Similarity 77.8%; Pred. No. 1.1;

Matches 14; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 KKLKKPPPLKKLKL 20  
||| |  
Db 1 KKLKKLKKLKKLKL 18

Search completed: June 6, 2003, 13:23:39  
Job time : 31.5 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 6, 2003, 13:22:11 ; Search time 12.5 Seconds  
(without alignments)  
153.815 Million cell updates/sec

Title: us-10-081-418-2  
Perfect score: 105  
Sequence: 1 KKKLLKKPPPLKKLLKKL 20

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues  
Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0  
Maximum DB seq length: 200000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	51	48.6	369	T24797	hypothetical prote
2	50	47.6	900	T47732	probable translati
3	50	47.6	900	T50773	translaton initia
4	49	46.7	110	D72863	Acort-107 protein
5	49	46.7	340	E81331	uroporphyrinogen d
6	49	46.7	656	S69554	hypothetical prote
7	49	46.7	1235	S24109	phosphorylase kina
8	48.5	46.2	668	C24785	hypothetical prote
9	48.5	46.2	700	C1H0H2	calpain (BC 3.4.22
10	48	45.7	700	S70028	mitogen-activated
11	48	45.7	339	E71918	uroporphyrinogen d
12	48	45.7	340	D64595	uroporphyrinogen d
13	48	45.7	867	A81707	ATP-dependent Clp
14	48	45.7	2897	B48666	cell proliferation
15	48	45.7	3256	A48666	cell proliferation
16	47	44.8	1324	S06187	RNA2 polypeptid -
17	47	44.8	85	D69386	conserved hypotet
18	46	43.8	157	S49448	oleosin - rape
19	46	43.8	537	F0MWV7	gag polyprotein -
20	46	43.8	1341	S66835	probable membrane
21	45.5	43.3	355	B75195	hypothetical prote
22	45.5	43.3	479	A64117	serine-type D-Ala-
23	45.5	43.3	700	S38361	calpain (EC 3.4.22
24	45	42.9	151	S07883	hypothetical prote
25	45	42.9	165	S50195	oleosin - rape
26	45	42.9	375	S58484	gag protein - maiz
27	45	42.9	493	A85433	sugar transporter
28	45	42.9	563	T40705	conserved hypotet
29	45	42.9	1152	H88533	probable ATP-depen

30	44.5	42.4	629	2	A71023	arginine-tRNA ligase
31	44	41.9	91	2	T08131	oleosin-like prote
32	44	41.9	143	2	JC4203	nitrogenase NifX c
33	44	41.9	143	2	T09232	nifX protein - Fra
34	44	41.9	186	2	T45838	hypothetical prote
35	44	41.9	224	2	F96668	protein FIN19.5 (1
36	44	41.9	319	2	S34806	hypothetical prote
37	44	41.9	469	2	JC5798	FOF1-ATPase (EC 3.
38	44	41.9	592	2	C04642	purh bifunctional
39	44	41.9	681	2	C84612	hypothetical prote
40	44	41.9	779	2	S56245	cell division cont
41	44	41.9	781	2	JC7382	DNA-directed DNA p
42	44	41.9	879	2	A56277	DNA-directed DNA p
43	44	41.9	893	2	F72253	hypothetical prote
44	44	41.9	1073	1	T08328	plasmid replicatio
45	44	41.9	1309	1	S35484	peptidyl-dipectida

ALIGNMENTS

RESULT 1  
T24797  
hypothetical protein T10C6.7 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 20-Jun-2000  
C:Accession: T24797  
R:McMurray, A.  
submitted to the EMBL Data Library, March 1997  
A:Reference number: Z19936  
A:Accession: T24797  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-369 <WILD>  
A:Cross-references: EMBL:Z93388; PIDN:GAB07663.1; GSPDB:GN00023; CESP:T10C6.7  
A:Experimental source: clone T10C6  
C:Genetics:  
A:Gene: CESP:T10C6.7  
A:Map position: 5  
A:Introns: 44/2; 114/2; 219/3; 313/2  
C:Superfamily: Caenorhabditis elegans hypothetical protein F08F3.6

Query Match 48.6% Score 51; DB 2; Length 369;  
Best Local Similarity 50.0%; Pred. No. 11;  
Matches 10; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

OY 1 KKKLLKKPPPLKKLLKKL 20  
DB 20 KKKRLQLPPPAFANILKIL 39

RESULT 2  
T47732  
probable translation initiation factor eIF-3 chain 8 F18021.110 [Imported] - Arabidops  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 20-Apr-2000 #sequence\_revision 20-Apr-2000 #text\_change 05-May-2000  
C:Accession: T47732  
R:Benes, V.; Warmbach, E.; Drzonek, H.; Ansoerge, W.; Mewes, H.W.; Rudd, S.; Lemcke, J.  
submitted to the Protein Sequence Database, April 2000  
A:Reference number: Z24474  
A:Accession: T47732  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-900 <BEN>  
A:Cross-references: EMBL:ATF18021  
A:Experimental source: cultivar Columbia; BAC clone F18021  
C:Genetics:  
A:Map position: 3  
A:Introns: 8/3; 593/3  
A:Note: F18021.110

Query Match 47.6% Score 50; DB 2; Length 900;  
Best Local Similarity 50.0%; Pred. No. 37;

Matches 9; Conservative 4; Mismatches 5; Indels 0; Gaps 0;  
 Oy 2 WKLLKKRPPLKKLKK 19  
 Db 215 WEKMLSKDKLKKLKK 232

## RESULT 3

T50773  
 translation initiation factor homolog p105 [imported] - Arabidopsis thaliana  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C:Date: 21-Jul-2000 #sequence\_revision 21-Jul-2000 #text\_change 21-Jul-2000  
 C:Accession: T50773  
 R:Karniol, B.; Yabaiom, A.; Krok, S.; Tsuge, T.; Matsui, M.; Deng, X.W.; Chamovitz, D.A.  
 FEBS Lett. 439, 173-179, 1998  
 A:Title: The Arabidopsis homolog of an eIF3 complex subunit associates with the COP9  
 A:Reference number: Z24448; MUID:9849901; PMID:9849901  
 A:Accession: T50773  
 A:Status: Preliminary;  
 A:Molecule type: mRNA  
 A:Residues: 1-900 <RAB>  
 A:Cross-references: EMBL:AF040102; PIDN:AAC83464.1

Query Match 47.6%; Score 49; DB 2; Length 900;  
 Best Local Similarity 50.0%; Pred. No. 37;  
 Matches 9; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

Oy 2 WKLLKKRPPLKKLKK 19  
 Db 215 WEKMLSKDKLKKLKK 232

## RESULT 4

D72863  
 AcOrf-107 protein - Autographa californica nuclear polyhedrosis virus  
 C:Species: Autographa californica nuclear polyhedrosis virus, ACMPV  
 A:Note: dsDNA virus  
 C:Date: 12-Nov-1999 #sequence\_revision 12-Nov-1999 #text\_change 12-Nov-1999  
 C:Accession: D72863  
 R:AYres, M.D.; Howard, S.C.; Kuzio, J.; Lopez-Ferber, M.; Possee, R.D.  
 Virology 202, 586-605, 1994  
 A:Title: The complete DNA sequence of Autographa californica nuclear polyhedrosis virus.  
 A:Reference number: A72850; MUID:94303173; PMID:8030224  
 A:Accession: D72863  
 A:Status: Preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-110 <AYR>  
 A:Cross-references: GB:L22858; NID:9510708; PIDN:AAA66737.1; PID:9559176  
 C:Genetics:  
 A:Gene: AcOrf-107

Query Match 46.7%; Score 49; DB 2; Length 110;  
 Best Local Similarity 64.7%; Pred. No. 5.7;  
 Matches 11; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

Oy 3 KKLKKRPPLKKLKK 19  
 Db 85 KCLLPAPPPOLKKLKK 101

## RESULT 5

E81331  
 uroporphyrinogen decarboxylase (EC 4.1.1.37) Cj1243 [imported] - Campylobacter jejuni (s  
 C:Species: Campylobacter jejuni  
 C:Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 03-Jun-2002  
 C:Accession: E81331  
 R:Parhill, J.; Wren, B.W.; Mungall, K.; Kelley, J.M.; Churche, C.; Basham, D.; Chitt  
 Nature 403, 665-668, 2000  
 A:Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals h  
 A:Reference number: AB1250; MUID:20150912; PMID:10688204  
 A:Accession: E81331  
 A:Status: preliminary

A:Molecule type: DNA  
 A:Residues: 1-340 <PAR>  
 A:Cross-references: GB:AL139077; GB:AL111168; NID:96968444; PIDN:CAB73497.1; PID:9696  
 A:Experimental source: serotype O2, strain NCTC 11168  
 C:Genetics:  
 A:Gene: heme; Cj1243  
 C:Superfamily: uroporphyrinogen decarboxylase  
 C:Keywords: carbon-carbon lyase; carboxy-lyase

Query Match 46.7%; Score 49; DB 2; Length 340;  
 Best Local Similarity 55.0%; Pred. No. 19;  
 Matches 11; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

Oy 1 KKKLKKRPPLKKLKK 20  
 Db 160 KCKKMLYQNPDLKKLKK 179

## RESULT 6

S69554  
 hypothetical protein YDR496c - yeast (Saccharomyces cerevisiae)  
 C:Species: Saccharomyces cerevisiae  
 C:Date: 22-Aug-1996 #sequence\_revision 06-Sep-1996 #text\_change 19-Apr-2002  
 C:Accession: S69554  
 R:Dietrich, F.S.  
 submitted to the EMBL Data Library, August 1995  
 A:Description: The sequence of S. cerevisiae cosmid 8166, 9787, 9717, and lambda 307  
 A:Reference number: S69553  
 A:Accession: S69554  
 A:Molecule type: DNA  
 A:Residues: 1-656 <DIE>  
 A:Cross-references: EMBL:U33057; NID:9927764; PID:9927766; GSPDB:GN00004; MIPS:YDR496  
 C:Genetics:  
 A:Gene: SGD:PUF6; MIPS:YDR496c  
 A:Cross-references: SGD:S0002904  
 A:Map position: 4R

Query Match 46.7%; Score 49; DB 2; Length 656;  
 Best Local Similarity 47.4%; Pred. No. 37;  
 Matches 9; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

Oy 2 WKLLKKRPPLKKLKK 20  
 Db 131 WERLKVKTPLPKQIREKL 149

## RESULT 7

S24109  
 phosphorylase kinase (EC 2.7.1.38) alpha chain, hepatic - rabbit  
 C:Species: Oryctolagus cuniculus (domestic rabbit)  
 C:Date: 25-Feb-1994 #sequence\_revision 10-Nov-1995 #text\_change 05-Nov-1999  
 C:Accession: S24109  
 R:Davidson, J.J.; Oezcelik, T.; Hamacher, C.; Williams, P.J.; Francke, U.; Killmann, M  
 Proc. Natl. Acad. Sci. U.S.A. 89, 2096-2100, 1992  
 A:Title: cDNA cloning of a liver isoform of the phosphorylase kinase alpha subunit an  
 A:Reference number: S24109; MUID:92196064; PMID:1372435  
 A:Accession: S24109  
 A:Status: Preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-1235 <DAV>  
 A:Cross-references: EMBL:X60421; NID:91657; PIDN:CMA42952.1; PID:91658  
 C:Keywords: liver; phosphotransferase

Query Match 46.7%; Score 49; DB 2; Length 1235;  
 Best Local Similarity 56.2%; Pred. No. 71;  
 Matches 9; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

Oy 5 LKKRPPLKKLKK 20  
 Db 734 LVDSFQPLKKLKK 749

## RESULT 8

C24785  
 Hypothetical protein 668 - slime mold (Dictyostelium discoideum) transposon DIRS-1 (frag  
 C:Species: Dictyostelium discoideum  
 C>Date: 23-Aug-1987 #sequence\_revision 23-Aug-1987 #text\_change 18-Jun-1993  
 C:Accession: C24785  
 R:Capello, J.; Handelsman, K.; Lodish, H.F.  
 Cell 43, 105-115, 1985  
 A:Title: Sequence of Dictyostelium DIRS-1: an apparent retrotransposon with inverted ter  
 A:Reference number: A94654; MUID:86079481; PMID:2416457  
 A:Accession: C24785  
 A:Molecule type: DNA  
 A:Residues: 1-668 <CAP>

Query Match 46.2%; Score 48.5; DB 2; Length 668;  
 Best Local Similarity 47.6%; Pred. No. 44;  
 Matches 10; Conservative 5; Mismatches 5; Indels 1; Gaps 1;  
 Oy 1 KMKLLKKPP-LLKKLLKL 20  
 Db 572 QMKCLAFPPILPILPILK 592

## RESULT 9

C1H0H2

calpain (EC 3.4.22.17) large chain 2 - human  
 N:Alternate names: calpain chain L-2; calpain II catalytic chain; high-calcium requiring  
 C:Species: Homo sapiens (man)  
 C>Date: 21-Nov-1993 #sequence\_revision 09-Aug-1997 #text\_change 16-Jul-1999  
 C:Accession: S10590; A31218; A33529  
 R:Sortimachi, H.; Ohmi, S.; Emori, Y.; Kawasaki, H.; Saito, T.C.; Ohno, S.; Minami, Y.; S  
 Biol. Chem. Hoppe-Seyler 371(Suppl.), 171-176, 1990  
 A:Title: A novel member of the calcium-dependent cysteine protease family.  
 A:Reference number: S10589; MUID:90380278; PMID:2400579  
 A:Accession: S10590  
 A:Molecule type: mRNA  
 A:Residues: 1-700 <SOR>  
 R:Imajoh, S.; Aoki, K.; Ohno, S.; Emori, Y.; Kawasaki, H.; Sugihara, H.; Suzuki, K.  
 Biochemistry 27, 8122-8128, 1988  
 A:Title: Molecular cloning of the cDNA for the large subunit of the high-Ca(2+)-regulin  
 A:Reference number: A31218; MUID:89166474; PMID:2852952  
 A:Accession: A31218  
 A:Molecule type: mRNA; protein  
 A:Residues: 1-210, 'I', 212-394, 'D', 396-445, 'I', 447-700 <TMA>  
 A:Cross-references: GB:M23254; NID:9511636; PIDN:AAA55645.1; PID:9511637  
 A:Note: parts of this sequence were determined by protein sequencing; the amino end of t  
 J. Biol. Chem. 264, 6404-6411, 1989  
 A:Title: Tandemly reiterated negative enhancer-like elements regulate transcription of a  
 A:Reference number: A33529; MUID:89197947; PMID:2539381  
 A:Accession: A33529  
 A:Molecule type: DNA  
 A:Residues: 1-67, 'G', 69-72, 'T', 75-78, 'R' <HAT>  
 A:Cross-references: DDBJ:J04700; NID:9179910; PIDN:AAA52760.1; PID:9463086  
 C:Genetics: GDB:CAPN2; mCAPN; CAPNml  
 A:Gene: GDB:CAPN2; mCAPN; CAPNml  
 A:Cross-references: GDB:119750; OMIM:114230  
 A:Map position: 1pter-1qter  
 C:Complex: heterodimer of L (large) and S (small) chains  
 C:Function:

A:Description: catalyzes the hydrolysis of peptides  
 A:Note: Cleaves preferentially after tyrosine, methionine, or arginine residues and bef  
 C:Superfamily: calpain large chain; calmodulin repeat homology; calpain catalytic domain  
 C:Keywords: acetylated amino end; calcium binding; cysteine proteinase; EF hand; heterod  
 F:2-700/Product: calpain large chain 2 #status predicted <CAP>  
 F:75-327/Domain: calpain catalytic domain homology <CAP>  
 F:529-560/Domain: calmodulin repeat homology <EF1>  
 F:572-604/Domain: calmodulin repeat homology <EF2>  
 F:605-634/Domain: calmodulin repeat homology <EF3>  
 F:637-668/Domain: calmodulin repeat homology <EF4>  
 F:2/Modified site: acetylated amino end (Ala) (in mature form) #status predicted  
 F:105,266/Active site: Cys, His, Asn #status predicted

Query Match 46.2%; Score 48.5; DB 1; Length 700;

Best Local Similarity 52.6%; Pred. No. 46;  
 Matches 10; Conservative 5; Mismatches 3; Indels 1; Gaps 1;  
 Oy 1 KMKLLKKPP-LLKKLLKL 19  
 Db 213 EWE-LKKPPNLEKTIK 230

## RESULT 10

S70028

mitogen-activated protein kinase kinase 1 - common carp (fragments)  
 C:Species: Cyprinus carpio (common carp)  
 C>Date: 14-Apr-1998 #sequence\_revision 24-Apr-1998 #text\_change 17-Mar-1999  
 C:Accession: S70028  
 R:Lee, J.H.; Lee, M.S.; Chen, K.T.; Chang, G.D.; Chou, C.K.; Huang, C.J.  
 Biochim. Biophys. Acta 1306, 133-136, 1996  
 A:Title: Genomic structure of carp mitogen-activated protein kinase kinase 1 gene.  
 A:Reference number: S70028; MUID:96221282; PMID:8634328  
 A:Accession: S70028  
 A:Status: Preliminary; not compared with conceptual translation  
 A:Molecule type: DNA  
 A:Residues: 1-7;8-14;15-21;22-28;29-35;36-42;43-49;50-56;57-63;64-70 <LEU>

Query Match 45.7%; Score 48; DB 2; Length 70;  
 Best Local Similarity 43.8%; Pred. No. 4.9;  
 Matches 7; Conservative 7; Mismatches 2; Indels 0; Gaps 0;  
 Oy 5 LKKPPPLKKLLKL 20  
 Db 50 IYNEPPVVKCLMKKM 65

## RESULT 11

E71918

uroporphyrinogen decarboxylase - Helicobacter pylori (strain J99)  
 C:Species: Helicobacter pylori  
 A:Variety: strain J99  
 C>Date: 12-Feb-1999 #sequence\_revision 12-Feb-1999 #text\_change 22-Jun-1999  
 C:Accession: E71918  
 R:Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Dolg, P.C.; Smith, D  
 Nature 397, 176-180, 1999  
 A:Title: Genomic sequence comparison of two unrelated isolates of the human gastric p  
 A:Reference number: A71800; MUID:99120557; PMID:9923682  
 A:Accession: E71918  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-339 <ARN>  
 A:Cross-references: GB:AE001487; GB:AE001439; NID:94155086; PIDN:AD06123.1; PID:9415  
 A:Experimental source: strain J99  
 C:Genetics:  
 A:Gene: hemE  
 C:Superfamily: uroporphyrinogen decarboxylase

Query Match 45.7%; Score 48; DB 2; Length 339;  
 Best Local Similarity 55.0%; Pred. No. 26;  
 Matches 11; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

Oy 1 KMKLLKKPP-LLKKLLKL 20  
 Db 159 KSKKMLSEPEVKALKEKL 178

## RESULT 12

D64595

uroporphyrinogen decarboxylase - Helicobacter pylori (strain 26695)  
 C:Species: Helicobacter pylori  
 C>Date: 09-Aug-1997 #sequence\_revision 09-Aug-1997 #text\_change 22-Jun-1999  
 C:Accession: D64595  
 R:Tombl, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R  
 Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khakh, H.G.; Glodek, A.; McK  
 son, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Matthey,  
 Nature 386, 539-547, 1997

A;Cross-references: EMBL:X65551  
C;Superfamily: kinase interaction domain homology

Search completed: June 6, 2003, 13:25:28  
Job time : 14.5 secs

Search completed: June 6, 2003, 13:25:28  
Job time : 14.5 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein. - protein search, using sw model

Run on: June 6, 2003, 13:22:41 ; Search time 15.5 Seconds  
(without alignments)  
133.213 Million cell updates/sec

Title: US-10-081-418-2

Perfect score: 105

Sequence: 1 KKKLKKPPPLKKLKL 20

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 392085 segs, 103240269 residues

Total number of hits satisfying chosen parameters: 392085

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications, AA.\*  
1: /cgn2\_6/ptodata/2/pubppaa/US08\_NEW\_PUB pep.\*  
2: /cgn2\_6/ptodata/2/pubppaa/PCT\_NEW\_PUB pep.\*  
3: /cgn2\_6/ptodata/2/pubppaa/US06\_NEW\_PUB pep.\*  
4: /cgn2\_6/ptodata/2/pubppaa/US06\_PUBCOMB pep.\*  
5: /cgn2\_6/ptodata/2/pubppaa/US07\_NEW\_PUB pep.\*  
6: /cgn2\_6/ptodata/2/pubppaa/US07\_PUBCOMB pep.\*  
7: /cgn2\_6/ptodata/2/pubppaa/PCTUS\_PUBCOMB pep.\*  
8: /cgn2\_6/ptodata/2/pubppaa/US08\_PUBCOMB pep.\*  
9: /cgn2\_6/ptodata/2/pubppaa/US09\_NEW\_PUB pep.\*  
10: /cgn2\_6/ptodata/2/pubppaa/US09\_PUBCOMB pep.\*  
11: /cgn2\_6/ptodata/2/pubppaa/US10\_NEW\_PUB pep.\*  
12: /cgn2\_6/ptodata/2/pubppaa/US10\_PUBCOMB pep.\*  
13: /cgn2\_6/ptodata/2/pubppaa/US60\_NEW\_PUB pep.\*  
14: /cgn2\_6/ptodata/2/pubppaa/US60\_PUBCOMB pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	105	100.0	20	9	US-10-081-418-2
2	56.5	53.8	16	9	US-09-876-904A-15
3	56.5	53.8	16	9	US-09-876-904A-629
4	53	50.5	17	9	US-09-999-745-55
5	53	50.5	17	9	US-09-554-000-39
6	50	47.6	18	10	US-09-447-966-1
7	49	46.7	26	10	US-09-030-619-125
8	48.5	46.2	23	9	US-09-820-053A-9
9	48.5	46.2	23	9	US-09-820-053A-56
10	48.5	46.2	700	9	US-10-038-557A-9
11	48.5	46.2	700	10	US-08-840-707A-9
12	48.5	46.2	700	10	US-09-768-877-23
13	48	45.7	513	9	US-10-051-902-20
14	48	45.7	513	9	US-10-051-909-20
15	46	43.8	3256	10	US-09-919-172-98
16	45	42.9	47	10	US-10-036-869-11
17	45	42.9	49	10	US-09-864-761-42909
18	45	42.9	49	10	US-09-864-761-40695
19	45	42.9	54	10	US-09-864-761-48671

20	45	42.9	74	10	US-09-864-761-33838	Sequence 33838, A
21	45	42.9	510	9	US-10-051-902-22	Sequence 22, Appl
22	45	42.9	510	9	US-10-051-909-22	Sequence 22, Appl
23	44.5	42.4	20	9	US-09-820-053A-68	Sequence 68, Appl
24	44.5	42.4	610	9	US-09-738-626-4495	Sequence 4495, Ap
25	44	41.9	22	9	US-09-820-053A-46	Sequence 46, Appl
26	44	41.9	22	9	US-09-820-053A-46	Sequence 46, Appl
27	44	41.9	22	9	US-09-820-053A-48	Sequence 48, Appl
28	44	41.9	22	9	US-09-820-053A-48	Sequence 48, Appl
29	44	41.9	22	10	US-09-764-853-157	Sequence 17, Appl
30	44	41.9	529	9	US-10-051-902-28	Sequence 609, App
31	44	41.9	529	9	US-10-051-902-28	Sequence 28, Appl
32	44	41.9	529	9	US-10-051-902-28	Sequence 28, Appl
33	44	41.9	529	9	US-10-051-902-28	Sequence 28, Appl
34	44	41.9	529	9	US-10-051-902-28	Sequence 28, Appl
35	44	41.9	529	9	US-10-051-902-28	Sequence 28, Appl
36	44	41.9	529	9	US-10-051-902-28	Sequence 28, Appl
37	43	41.0	23	9	US-09-954-043-4	Sequence 29, Appl
38	43	41.0	23	9	US-10-102-806-678	Sequence 4, Appl1
39	43	41.0	23	9	US-09-820-053A-9	Sequence 678, App
40	43	41.0	26	9	US-09-820-053A-161	Sequence 161, App
41	43	41.0	26	10	US-09-908-1395-2	Sequence 122, App
42	43	41.0	216	12	US-10-057-510-4	Sequence 4, Appl1
43	43	41.0	242	10	US-09-764-898-199	Sequence 199, App
44	43	41.0	489	10	US-09-956-425-6	Sequence 6, Appl1
45	43	41.0	491	9	US-09-029-327-2	Sequence 1, Appl1
			491	10	US-09-888-077-1	

## ALIGNMENTS

RESULT 1  
US-10-081-418-2  
Sequence 2, Application US/10081418  
Publication No. US20030096745A1  
GENERAL INFORMATION:  
APPLICANT: HAHM, Kyung-Soo  
APPLICANT: LEE, Dong Gun  
APPLICANT: PARK, Yoonkyung  
APPLICANT: KIM, Hee Nam  
TITLE OF INVENTION: NO. US20030096745A1 peptides with increased + charge and hydrophobicity by substituting one or more amino acids of CA-M peptide and  
TITLE OF INVENTION: Pharmaceutical compositions containing thereof  
FILE REFERENCE: 428.1014  
CURRENT APPLICATION NUMBER: US/10/081,418  
CURRENT FILING DATE: 2002-02-22  
NUMBER OF SEQ ID NOS: 2  
SOFTWARE: Kopatentia 1.71  
SEQ ID NO 2  
LENGTH: 20  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: peptide with increased + charge and hydrophobicity by substituting amino acids of SEQ. ID. NO 1 with lysine and leuc  
US-10-081-418-2  
Query Match 100.0%; Score 105; DB 9; Length 20;  
Best Local Similarity 100.0%; Pred. No. 1.7e-07;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 KKKLKKPPPLKKLKL 20  
DB 1 KKKLKKPPPLKKLKL 20  
RESULT 2  
US-09-876-904A-15  
Sequence 15, Application US/09876904A  
Publication No. US20030072794A1  
GENERAL INFORMATION:  
APPLICANT: BOULIKAS, TENI  
TITLE OF INVENTION: ENCAPSULATION OF PLASMID DNA (LIPOGENES TM) AND THERAPEUTIC

;; TITLE OF INVENTION: AGENTS WITH NUCLEAR LOCALIZATION SIGNAL/FUSOGENIC PEPTIDE  
;; TITLE OF INVENTION: CONJUGATES INTO TARGETED LIPOSOME COMPLEXES  
;; FILE REFERENCE: TB-2002.00  
;; CURRENT APPLICATION NUMBER: US/09/876,904A  
;; CURRENT FILING DATE: 2001-06-08  
;; PRIOR APPLICATION NUMBER: US 60/210,925  
;; PRIOR FILING DATE: 2000-06-09  
;; NUMBER OF SEQ ID NOS: 629  
;; SOFTWARE: PatentIn Ver. 2.1  
;; SEQ ID NO 15  
;; LENGTH: 16  
;; TYPE: PRT  
;; ORGANISM: Artificial Sequence  
;; FEATURE:  
;; OTHER INFORMATION: Description of Artificial Sequence: Synthetic amphiphilic  
;; OTHER INFORMATION: fusogenic peptide  
US-09-876-904A-15

Query Match 53.8%; Score 56.5; DB 9; Length 16;  
Best Local Similarity 83.3%; Pred. No. 0.25;  
Matches 15; Conservative 0; Mismatches 0; Indels 3; Gaps 1;

OY 3 KKLKKPPPLKKLKL 20  
DB 2 KKLKK---LKKLKLKL 16

RESULT 3  
US-09-876-904A-629  
;; Sequence 629, Application US/09876904A  
;; Publication No. US20030072794A1  
;; GENERAL INFORMATION:  
;; APPLICANT: BOULIKAS, TENI  
;; TITLE OF INVENTION: ENCAPSULATION OF PLASMID DNA (LIPOGENES TM) AND THERAPEUTIC  
;; TITLE OF INVENTION: AGENTS WITH NUCLEAR LOCALIZATION SIGNAL/FUSOGENIC PEPTIDE  
;; FILE REFERENCE: TB-2002.00  
;; CURRENT APPLICATION NUMBER: US/09/876,904A  
;; CURRENT FILING DATE: 2001-06-08  
;; PRIOR APPLICATION NUMBER: US 60/210,925  
;; PRIOR FILING DATE: 2000-06-09  
;; NUMBER OF SEQ ID NOS: 629  
;; SOFTWARE: PatentIn Ver. 2.1  
;; SEQ ID NO 629  
;; LENGTH: 16  
;; TYPE: PRT  
;; ORGANISM: Artificial Sequence  
;; FEATURE:  
;; OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide  
US-09-876-904A-629

Query Match 53.8%; Score 56.5; DB 9; Length 16;  
Best Local Similarity 83.3%; Pred. No. 0.25;  
Matches 15; Conservative 0; Mismatches 0; Indels 3; Gaps 1;

OY 3 KKLKKPPPLKKLKL 20  
DB 2 KKLKK---LKKLKLKL 16

RESULT 4  
US-09-999-745-55  
;; Sequence 55, Application US/09999745  
;; Patent No. US20020157120A1  
;; GENERAL INFORMATION:  
;; APPLICANT: THE REGENTS OF THE UNIVERSITY OF CALIFORNIA  
;; APPLICANT: Tsien, Roger Y.  
;; APPLICANT: Baird, Geoffrey  
;; TITLE OF INVENTION: CIRCULARLY PERMUTED FLUORESCENT PROTEIN INDICATORS  
;; FILE REFERENCE: REGEN1470-1  
;; CURRENT APPLICATION NUMBER: US/09/999,745  
;; CURRENT FILING DATE: 2001-10-23  
;; PRIOR APPLICATION NUMBER: 09/316,920

;; PRIOR FILING DATE: 1999-05-21  
;; NUMBER OF SEQ ID NOS: 67  
;; SOFTWARE: PatentIn version 3.0  
;; SEQ ID NO 55  
;; LENGTH: 17  
;; TYPE: PRT  
;; ORGANISM: Artificial Sequence  
;; FEATURE:  
;; OTHER INFORMATION: Calmodulin binding peptide-2  
US-09-999-745-55

Query Match 50.5%; Score 53; DB 9; Length 17;  
Best Local Similarity 76.5%; Pred. No. 0.75;  
Matches 13; Conservative 0; Mismatches 0; Indels 4; Gaps 1;

OY 2 WKLLKKPPPLKKLKL 18  
DB 3 WKLLKK---LKKLKLKL 15

RESULT 5  
US-09-554-000-39  
;; Sequence 39, Application US/09554000  
;; Patent No. US20020165364A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Tsien, Roger Y.  
;; APPLICANT: Miyawaki, Atsushi  
;; TITLE OF INVENTION: FLUORESCENT PROTEIN SENSORS FOR  
;; TITLE OF INVENTION: DETECTION OF ANALYTES  
;; FILE REFERENCE: 07257/042001  
;; CURRENT APPLICATION NUMBER: US/09/554,000  
;; CURRENT FILING DATE: 2000-04-20  
;; PRIOR APPLICATION NUMBER: 08/818,252  
;; PRIOR FILING DATE: 1997-03-14  
;; NUMBER OF SEQ ID NOS: 56  
;; SOFTWARE: FastSeq for Windows Version 4.0  
;; SEQ ID NO 39  
;; LENGTH: 17  
;; TYPE: PRT  
;; ORGANISM: Artificial Sequence  
;; FEATURE:  
;; OTHER INFORMATION: Calmodulin binding peptide-2  
US-09-554-000-39

Query Match 50.5%; Score 53; DB 9; Length 17;  
Best Local Similarity 76.5%; Pred. No. 0.75;  
Matches 13; Conservative 0; Mismatches 0; Indels 4; Gaps 1;

OY 2 WKLLKKPPPLKKLKL 18  
DB 3 WKLLKK---LKKLKLKL 15

RESULT 6  
US-09-447-966-1  
;; Sequence 1, Application US/09447966  
;; Patent No. US20010004636A1  
;; GENERAL INFORMATION:  
;; APPLICANT: WOLFF, JON A  
;; APPLICANT: Monahan, Sean D  
;; APPLICANT: Slatum, Paul M E  
;; APPLICANT: Hagstrom, James E  
;; APPLICANT: Budker, Vladimir G  
;; APPLICANT: Roizema, David B  
;; TITLE OF INVENTION: INTRAVASCULAR DELIVERY OF NON-VIRAL NUCLEIC ACID  
;; FILE REFERENCE: Mirus.013.03  
;; CURRENT APPLICATION NUMBER: US/09/447,966  
;; CURRENT FILING DATE: 1999-11-23  
;; PRIOR APPLICATION NUMBER: 60/121730  
;; PRIOR FILING DATE: 1999-02-26  
;; PRIOR APPLICATION NUMBER: 60/146564  
;; PRIOR FILING DATE: 1999-07-30  
;; NUMBER OF SEQ ID NOS: 1



```

; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 1
; LENGTH: 18
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: 18-mer
US-09-447-966-1

Query Match
Best Local Similarity 47.6%; Score 50; DB 10; Length 18;
Matches 13; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 4 KLLKKPPPLKKLKL 20
Db 1 KLLKKLKLKKLKKLKL 17

RESULT 7
US-09-030-619-125
; Sequence 125, Application US/090306195
; Patent No. US20020035061A1
; GENERAL INFORMATION:
; APPLICANT: Krieger, Timothy J.
; APPLICANT: Taylor, Robert
; APPLICANT: Erfle, Douglas
; APPLICANT: Fraser, Janet R.
; APPLICANT: West, Michael H.P.
; APPLICANT: McNicol, Patricia J.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATING
; TITLE OF INVENTION: INFECTIONS USING CATIONIC PEPTIDES ALONE OR IN COMBINATION
; FILE REFERENCE: 660081.406
; CURRENT APPLICATION NUMBER: US/09/030,619B
; CURRENT FILING DATE: 1998-02-25
; NUMBER OF SEQ ID NOS: 232
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 125
; LENGTH: 26
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Cationic Peptide Analogue
US-09-030-619-125

Query Match
Best Local Similarity 46.7%; Score 49; DB 10; Length 26;
Matches 9; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 1 KKKKLLKKPPPLKKLKL 17
Db 1 KKKLPLKKLTPAVKKVL 17

RESULT 8
US-09-820-053A-49
; Sequence 49, Application US/09820053A
; Publication No. US20030083243A1
; GENERAL INFORMATION:
; APPLICANT: Owen, Donald R.
; TITLE OF INVENTION: SHORT BIOACTIVE PEPTIDES
; FILE REFERENCE: HELX027
; CURRENT APPLICATION NUMBER: US/09/820,053A
; CURRENT FILING DATE: 2001-03-28
; NUMBER OF SEQ ID NOS: 165
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 49
; LENGTH: 23
; TYPE: PRT
; ORGANISM: ARTIFICIAL SEQUENCE
; FEATURE:
; OTHER INFORMATION: SYNTHETIC SEQUENCE
```

```

; NAME/KEY: MOD_RES
; LOCATION: (23)
; OTHER INFORMATION: AMIDATION
US-09-820-053A-49

Query Match
Best Local Similarity 46.2%; Score 48.5; DB 9; Length 23;
Matches 12; Conservative 0; Mismatches 7; Indels 1; Gaps 1;

QY 1 KKKKLLKKPPPLKKLKL 20
Db 1 KW-KLKKKTKLKKKFAKRL 19

RESULT 9
US-09-820-053A-56
; Sequence 56, Application US/09820053A
; Publication No. US20030083243A1
; GENERAL INFORMATION:
; APPLICANT: Owen, Donald R.
; TITLE OF INVENTION: SHORT BIOACTIVE PEPTIDES
; FILE REFERENCE: HELX027
; CURRENT APPLICATION NUMBER: US/09/820,053A
; CURRENT FILING DATE: 2001-03-28
; NUMBER OF SEQ ID NOS: 165
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 56
; LENGTH: 23
; TYPE: PRT
; ORGANISM: ARTIFICIAL SEQUENCE
; FEATURE:
; OTHER INFORMATION: SYNTHETIC SEQUENCE
US-09-820-053A-56

Query Match
Best Local Similarity 46.2%; Score 48.5; DB 9; Length 23;
Matches 12; Conservative 0; Mismatches 7; Indels 1; Gaps 1;

QY 1 KKKKLLKKPPPLKKLKL 20
Db 1 KW-KLKKKTKLKKKFAKRL 19

RESULT 10
US-10-038-557A-9
; Sequence 9, Application US/10038557A
; Publication No. US20030092684A1
; GENERAL INFORMATION:
; APPLICANT: Fredelking, Terry M.
; APPLICANT: Ignatyev, George M.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATING HEMORRHAGIC VIRUS
; TITLE OF INVENTION: INFECTIONS AND OTHER DISORDERS
; FILE REFERENCE: 24881-301D
; CURRENT APPLICATION NUMBER: US/10/038,557A
; CURRENT FILING DATE: 2002-06-05
; PRIOR APPLICATION NUMBER: 09/840,707
; PRIOR FILING DATE: 2001-04-23
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: 09/562,979
; PRIOR FILING DATE: 1999-04-27
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 9
; LENGTH: 700
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: IL-1 receptor intracellular ligand protein
; OTHER INFORMATION: comprising amino acid sequence
; PUBLICATION INFORMATION:
; PATENT DOCUMENT NUMBER: 5,817,476
; PATENT FILING DATE: 1995-06-07
```

PUBLICATION DATE: 1998-10-06  
US-10-038-557A-9

Query Match 46.2%; Score 48.5; DB 9; Length 700;  
Best Local Similarity 52.6%; Pred. No. 1.2e+02;  
Matches 10; Conservative 5; Mismatches 3; Indels 1; Gaps 1;

QY 1 KWKKLKKPPPLKKLKK 19  
Db 213 EWYE-LKKPPNLFKIIQK 230

RESULT 11  
US-09-840-707A-9  
Sequence 9, Application US/09840707A  
Patent No. US2002007276A1  
GENERAL INFORMATION:  
APPLICANT: Fredeking, Terry M.  
APPLICANT: Ignatyev, George M.  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATING HEMORRHAGIC VIRUS  
FILE REFERENCE: 24881-301C  
CURRENT APPLICATION NUMBER: US/09/840,707A  
CURRENT FILING DATE: 2001-04-23  
PRIOR APPLICATION NUMBER: 09/562,979  
PRIOR FILING DATE: 2000-04-27  
PRIOR APPLICATION NUMBER: 60/198,210  
PRIOR FILING DATE: 1999-04-27  
NUMBER OF SEQ ID NOS: 26  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 9  
LENGTH: 700  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: IL-1 receptor intracellular ligand protein  
OTHER INFORMATION: comprising amino acid sequence  
PUBLICATION INFORMATION:  
PATENT DOCUMENT NUMBER: 5,817,476  
PATENT FILING DATE: 1995-06-07  
PUBLICATION DATE: 1998-10-06  
US-09-840-707A-9

Query Match 46.2%; Score 48.5; DB 10; Length 700;  
Best Local Similarity 52.6%; Pred. No. 1.2e+02;  
Matches 10; Conservative 5; Mismatches 3; Indels 1; Gaps 1;

QY 1 KWKKLKKPPPLKKLKK 19  
Db 213 EWYE-LKKPPNLFKIIQK 230

RESULT 12  
US-09-768-877-23  
Sequence 23, Application US/09768877  
Patent No. US20020150896A1  
GENERAL INFORMATION:  
APPLICANT: POLONSKY, KENNETH S.  
APPLICANT: HORIKAWA, YUKIO  
APPLICANT: ODA, NAOHISA  
APPLICANT: COX, NANCY J.  
APPLICANT: SREENAN, SEAMUS  
APPLICANT: ZHOU, YUN-PING  
APPLICANT: OTANI, KENICHI  
APPLICANT: HANIS, CRAIG L.  
APPLICANT: BELL, GRAEME I.  
TITLE OF INVENTION: METHODS OF TREATMENT OF TYPE 2 DIABETES  
FILE REFERENCE: ARCD:307  
CURRENT APPLICATION NUMBER: US/09/768,877  
CURRENT FILING DATE: 2001-01-23  
PRIOR APPLICATION NUMBER: 09/422,869  
PRIOR FILING DATE: 1999-10-21  
NUMBER OF SEQ ID NOS: 30

SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 23  
LENGTH: 700  
TYPE: PRT  
ORGANISM: Human  
US-09-768-877-23

Query Match 46.2%; Score 48.5; DB 10; Length 700;  
Best Local Similarity 52.6%; Pred. No. 1.2e+02;  
Matches 10; Conservative 5; Mismatches 3; Indels 1; Gaps 1;

QY 1 KWKKLKKPPPLKKLKK 19  
Db 213 EWYE-LKKPPNLFKIIQK 230

RESULT 13  
US-10-051-902-20  
Sequence 20, Application US/10051902  
Patent No. US20020178468A1  
GENERAL INFORMATION:  
APPLICANT: Allen, Steve  
APPLICANT: Hitz, Bill  
APPLICANT: Kinney, Tony  
APPLICANT: Tinney, Scott  
TITLE OF INVENTION: Plant Sugar Transport Proteins  
FILE REFERENCE: BB-1163  
CURRENT APPLICATION NUMBER: US/10/051,902  
CURRENT FILING DATE: 2002-01-17  
PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: US/09/291,922  
PRIOR FILING DATE: EARLIER FILING DATE: 1999-04-14  
NUMBER OF SEQ ID NOS: 30  
SOFTWARE: Microsoft Office 97  
SEQ ID NO 20  
LENGTH: 513  
TYPE: PRT  
ORGANISM: Zea mays  
US-10-051-902-20

Query Match 45.7%; Score 48; DB 9; Length 513;  
Best Local Similarity 43.8%; Pred. No. 1e+02;  
Matches 7; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY 2 WKLLKKPPPLKKL 17  
Db 267 WKLLSPTPAVRRL 282

RESULT 14  
US-10-051-909-20  
Sequence 20, Application US/10051909  
Patent No. US20020199217A1  
GENERAL INFORMATION:  
APPLICANT: Allen, Steve  
APPLICANT: Helentjaris, Tim  
APPLICANT: Hitz, Bill  
APPLICANT: Kinney, Tony  
APPLICANT: Tinney, Scott  
TITLE OF INVENTION: Plant Sugar Transport Proteins  
FILE REFERENCE: BB1163 US CIP  
CURRENT APPLICATION NUMBER: US/10/051,909  
CURRENT FILING DATE: 2002-01-17  
PRIOR APPLICATION NUMBER: 60/083,044  
PRIOR FILING DATE: April 24, 1998  
NUMBER OF SEQ ID NOS: 38  
SOFTWARE: Microsoft Office 97  
SEQ ID NO 20  
LENGTH: 513  
TYPE: PRT  
ORGANISM: Zea mays  
US-10-051-909-20

Query Match 45.7%; Score 48; DB 9; Length 513;

Best Local Similarity 43.88; Pred. No. 1e+02;  
Matches 7; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

Matches	7	Conservative	5	Mismatches	4	Indels	0	Gaps	0
---------	---	--------------	---	------------	---	--------	---	------	---

QY 2 WKKLLKKPPPLKKLL 17

Db 267 WKKLILSPTPAVRIL 282

RESULT 15  
US-09-919-172-98

US-09-919-172-98

sequence 98, Application US/09919172  
Patent No. US20020119463A1

Patent No. US20020119463A1

GENERAL INFORMATION:

APPLICANT: Farls, Mary

APPLICANT: Turner, Christopher M.

**TITLE OF INVENTION: PROSTATE CANCER MARKERS**

FILE REFERENCE: PA-0036 US

CURRENT APPLICATION NUMBER: IIS/09/919 172

CURRENT FILING DATE: 2001-07

PRIOR APPLICATION NUMBER: 60/333 450  
CURRENT FILING DATE: 2001-07-30

PRIOR APPLICATION NUMBER: 60/

PRIOR FILING DATE: 200

NUMBER OF S...

; SOFTWARE: PER

; SEQ ID NO 98

LENGTH: 3256

TYPE: PR

ORGANISM: *Homo sapiens*

ORGANISM: Homo sapiens

NAME/REV: m100 features

OTHER INFORMATION: Incyte ID No. US20020119463A1 27001332CD1  
US-09-919-172-98

Query Match	45.78;	Score	48;	DB	10;	Length	3256;
Best Local Similarity	47.48;	Pred	NO		6	7e+02.	

Best local similarity	4/.44	Pred. No. 6.7e+02
Matches	9	Conservative 5; Mismatches 5; Indels 0; Gaps 0;

1 KWKKLLKKPPPLKKLLK 19

Db 535 KRKSLVMTPTPLKKIIE 553

Search completed: June 6, 2003, 13:26:03  
Job time : 16.5 secs

Job time : 16.5 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 6, 2003, 13:21:26 ; Search time 23.5 Seconds

(without alignments)  
175.359 Million cell updates/sec

Title: US-10-081-418-2

Perfect score: 105

Sequence: 1 KMKRLKKPPPLKRLKRL 20

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SPTREMBL.21.\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_virus:\*  
16: sp\_bacteriophage:\*  
17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	53	50.5	357	5	O9GRQ3
2	52.5	50.0	173	10	O9FYP3
3	51	48.6	341	16	O8XMH7
4	51	48.6	369	5	O45769
5	50	47.6	900	10	O9LWM6
6	50	47.6	1437	5	O9GQ51
7	49	46.7	340	16	O9PNS4
8	49	46.7	656	3	O04373
9	48.5	46.2	608	5	O96849
10	48.5	46.2	608	5	O23898
11	48.5	46.2	700	4	O9HBB1
12	48.5	46.2	700	4	O8WU26
13	48.5	46.2	700	6	O9G1G1
14	48	45.7	552	13	O9YGC1
15	48	45.7	867	16	O9PKS5
16	47	44.8	110	15	O9WCA9

17	47	44.8	215	15	O9VIF1	O9VIF1 human immun
18	47	44.8	369	10	O9PMQ2	O9PMQ2 arabidopsis
19	47	44.8	524	15	O9ORL7	O9ORL7 porcine end
20	46.5	44.3	588	5	O9VFL0	O9VFL0 drosophila
21	46	43.8	85	17	O29172	O29172 archaeoglob
22	46	43.8	157	10	O39338	O39338 brassica na
23	46	43.8	373	16	O8R956	O8R956 thermococcus
24	46	43.8	395	5	O9VD68	O9VD68 drosophila
25	46	43.8	529	10	O9FP52	O9FP52 oryza sativ
26	46	43.8	856	11	O9R2C5	O9R2C5 mus musculu
27	46	43.8	858	12	O9IBW9	O9IBW9 turkey herp
28	46	43.8	858	12	O9E6R1	O9E6R1 turkey herp
29	46	43.8	905	11	O9WUX5	O9WUX5 mus musculu
30	46	43.8	1341	3	O08281	O08281 saccharomyc
31	46	43.8	1342	3	O93271	O93271 saccharomyc
32	45.5	43.3	306	2	O93RK1	O93RK1 haemophilus
33	45.5	43.3	355	17	O9V219	O9V219 pyrococcus
34	45.5	43.3	421	11	O8R486	O8R486 mus musculu
35	45.5	43.3	424	2	O93RK2	O93RK2 haemophilus
36	45.5	43.3	896	5	O94891	O94891 drosophila
37	45	42.9	120	15	O9ODY5	O9ODY5 porcine end
38	45	42.9	165	10	O93352	O93352 brassica na
39	45	42.9	299	4	O99785	O99785 homo sapien
40	45	42.9	325	13	O9PVL2	O9PVL2 gallus gall
41	45	42.9	342	10	O9FHJ5	O9FHJ5 arabidopsis
42	45	42.9	345	10	O9XET2	O9XET2 zea mays (m
43	45	42.9	436	12	O9DVS6	O9DVS6 pluteella xy
44	45	42.9	485	5	O95RE0	O95RE0 drosophila
45	45	42.9	493	10	O23213	O23213 arabidopsis

## ALIGNMENTS

## RESULT 1

ID O9GRQ3 PRELIMINARY; PRT; 357 AA.  
AC O9GRQ3;  
DT 01-MAR-2001 (TREMBLrel. 16, Created)  
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)  
DE Hypothetical 39.6 kDa protein.  
GN L6520.11.  
OS Leishmania major.  
OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.  
OX NCBI\_TaxID=5664;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=FRIEDLIN;  
RA Masny D., Purnelle B., Goffeau A., Ivens A.C., Quall M.,  
RA Rajandream M.A., Barrell B.G.;  
RL Submitted (Oct-2000) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=FRIEDLIN;  
RX MEDLINE=98146435; PubMed=9477341;  
RA Ivens A.C., Lewis S.M., Bagherzadeh A., Zhang L., Chan H.M.,  
RA Smith D.F.;  
RT "A physical map of the Leishmania major Friedlin genome.";  
RL Genome Res. 8:135-145(1998).  
DR EMBL; AL446005; CAC14538.1; -  
KW Hypothetical protein.  
SQ SEQUENCE 357 AA; 39621 MW; D27C03EFP2B2139E CMC64;

Query Match 50.5%; Score 53; DB 5; Length 357;  
Best Local Similarity 62.5%; Pred. No. 9.1;  
Matches 10; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

OY 4 KMKPPPLKRLKRL 19  
DB 85 KMKPPPLKRLKRL 100

```

RESULT 2
O9FCJP3 PRELIMINARY; PRT: 173 AA.
ID O9FCJP3;
AC O9FCJP3;
DT 01-MAR-2001 (TReMBLrel. 16, Created)
DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)
DE 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE 50S ribosomal protein L29 (Putative 50S ribosomal protein L29).
AT5G65220.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702.
RN
RP SEQUENCE FROM N.A.
RC STRAIN-COLUMBIA;
RX MEDLINE=98403884; PubMed=9734815;
RA Kotani H., Nakamura Y., Sato S., Asamizu E., Kaneko T., Miyajima N.,
RA Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 5. VI.
RT Sequence features of the regions of 1,367,185 bp covered by 19
RT physically assigned P1 and TAC clones."
RL DNA Res. 5:203-216(1998).
RN
RN [2]
RP SEQUENCE FROM N.A.
RA Yamada K., Liu S.X., Pham P.K., Banh J., Dale J.M., Goldsmith A.D.,
RA Jiang P.X., Lee J.M., Onodera C.S., Quach H.L., Tang C., Toriumi M.,
RA Yamamura Y., Yu G., Yu S., Bowser L., Carninci P., Chen H., Cheuk R.,
RA Hayashizaki Y., Ishida J., Jones T., Kamiya A., Karlin-Neumann G.,
RA Kawai J., Kim C., Kosemura E., Lam B., Lin J., Meyers M.C., Miranda M.,
RA Narusaka M., Nguyen M., Palm C.J., Sakurai T., Satou M., Seki M.,
RA Shinn P., Southwick A., Tracy S.E., Shinozaki K., Davis R.W.,
RA Ecker J.R., Theologis A.;
RT "Full length cDNA of gene MON23.16/AT5G65220 (GI:10178184).";
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
RN
RN [3]
RP SEQUENCE FROM N.A.
RA Yamada K., Liu S.X., Sakano H., Pham P.K., Banh J., Etny P., Lee J.M.,
RA Toriumi M., Yu G., Brooks S., Chao O., Chen H., Karlin-Neumann G.,
RA Kim C., Lam B., Miranda M., Nguyen M., Palm C.J., Shinn P.,
RA Southwick A., Davis R.W., Ecker J.R., Theologis A.;
RT "Arabidopsis Open Reading Frame (ORF) Clones.";
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL: AB013395; BAB1658.1;
DR EMBL: AY045966; AAK76640.1;
DR EMBL: AY079351; AAL85082.1;
DR InterPro: IPR001854; Ribosomal_L29.
DR Pfam: PF00831; Ribosomal_L29; 1.
DR TIGRFAMs: TIGR00012; L29; 1.
KM Ribosomal protein.
SQ SEQUENCE 173 AA; 19377 MW; B21C46FC9364CE42 CRC64;

Query Match 50.0%; Score 52.5; DB 10; Length 173;
Best Local Similarity 58.8%; Pred. No. 5.6;
Matches 10; Conservative 4; Mismatches 2; Indels 1; Gaps 1;

OY 1 KMKK-LTKKPPPLKTL 16
DB 141 QMKSIYPRPPSLKTL 157

```

```

OC Clostridiales; Clostridiaceae; Clostridium.
OX NCBI_TaxID=1502;
RN
RP SEQUENCE FROM N.A.
RC STRAIN-13 / TYPE A;
RX PubMed=11792842;
RA Shimizu T., Ohtani K., Hirakawa H., Ohshima K., Yamashita A.,
RA Shiba T., Ogasawara N., Hattori M., Kuhara S., Hayashi H.;
RT "Complete genome sequence of Clostridium perfringens, an anaerobic
RT flesh-eater.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:996-1001(2002).
DR EMBL: AP003187; BAB80418.1;
KM Hypothetical protein; Complete proteome.
SQ SEQUENCE 341 AA; 40334 MW; 196734ADC0F00C4E CRC64;

Query Match 48.6%; Score 51; DB 16; Length 341;
Best Local Similarity 58.8%; Pred. No. 17;
Matches 10; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

OY 3 KKLKKRPPLKTLKK 19
DB 5 KKLKKMPKMKSKFK 21

RESULT 4
O45769 PRELIMINARY; PRT: 369 AA.
ID O45769;
AC O45769;
DT 01-JUN-1998 (TReMBLrel. 06, Created)
DT 01-JUN-1998 (TReMBLrel. 06, Last sequence update)
DE 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE T10C6.7 protein.
GN T10C6.7.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN
RP SEQUENCE FROM N.A.
RA McMurray A.A.;
RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.
RN
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99065613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C. elegans: A platform for
RT investigating biology.";
RL Science 282:2012-2018(1998).
DR EMBL: Z93388; CAB07663.1;
DR InterPro: IPR001810; F-box.
DR Pfam: PF00646; F-box; 1.
DR PROSITE: PS50181; FBOX; 1.
SQ SEQUENCE 369 AA; 43506 MW; 00DEC66EAF291041 CRC64;

Query Match 48.6%; Score 51; DB 5; Length 369;
Best Local Similarity 50.0%; Pred. No. 18;
Matches 10; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

OY 1 KMKKLKKRPPLKTLKTL 20
DB 20 KFKRLQLPPPAFANILKIL 39

RESULT 5
O9LYM6 PRELIMINARY; PRT: 900 AA.
ID O9LYM6;
AC O9LYM6;
DT 01-OCT-2000 (TReMBLrel. 15, Created)
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Probable eukaryotic translation initiation factor 3 subunit 8
DE (Putative eukaryotic translation initiation factor 3 subunit 8).
GN F1802L110 OR AT3G56150.

```



RA Hunnicke-Smith S., Hyman R., Komp C., Lashkari D., Lew H., Lin D.,  
 RA Mosedale D., Nakahara K., Namath A., Oefner P., Oh C., Petel F.X.,  
 RA Robert D., Schramm S., Schroeder M., Shogren T., Shroff N.,  
 RA Wiant A., Yelton M., Bolstein D., Davis R.W.;  
 RL Submitted (AUG-1995) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Dietrich F.S.;  
 RL Submitted (AUG-1995) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA Jia Y., Cherry J.M.;  
 RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: U33057; AAB64938.1; -.  
 DR SGD: S0002904; YDR496C.  
 DR InterPro: IPR001313; Pumlilo/Puf.  
 DR Pfam: PF00806; PUF; 4.  
 DR SMART: SM00025; Pumlilo; 6.  
 SQ SEQUENCE 656 AA; 75105 MW; 4D172E8854326ECF CRC64;

Query Match 46.7%; Score 49; DB 3; Length 656;  
 Best Local Similarity 47.4%; Pred. No. 58;  
 Matches 9; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

Qy 2 WKRLKKPPPLKLLKL 20  
 Db 131 WKRLKKPPPLKLLKL 149

## RESULT 9

ID 096849 PRELIMINARY; PRT; 608 AA.  
 AC 096849;  
 DT 01-MAY-1999 (TREMblrel. 10, Created)  
 DT 01-MAY-1999 (TREMblrel. 10, Last sequence update)  
 DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)  
 DE Slime mold (D.discoidium) transposon DIRS-1, complete, clone  
 DE SB41.  
 OS Dictyostelium discoideum (Slime mold).  
 OC Eukaryota; Mycetozoa; Dictyostellida; Dictyostelium.  
 OX NCBI\_Taxid=44689;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-AX-3;  
 RX MEDLINE=85061212; PubMed=6095047;  
 RA Cappello J., Cohen S.M., Lodish H.F.;  
 RT "Dictyostelium transposable element DIRS-1 preferentially inserts into  
 RT DIRS-1 sequences.";  
 RL Mol. Cell. Biol. 4:2207-2213(1984).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-AX-3;  
 RX MEDLINE=85085937; PubMed=6096693;  
 RA Cohen S.M., Cappello J., Lodish H.F.;  
 RT "Transcription of Dictyostelium discoideum transposable element DIRS-  
 RT 1.";  
 RL Mol. Cell. Biol. 4:2332-2340(1984).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-AX-3;  
 RX MEDLINE=86079481; PubMed=2416457;  
 RA Cappello J., Handelsman K., Lodish H.F.;  
 RT "Sequence of Dictyostelium DIRS-1: An apparent retrotransposon with  
 RT inverted terminal repeats and an internal circle junction sequence.";  
 RL EMBL: M11339; AAA33195.1; -.  
 DR InterPro: IPR000477; RVTse.  
 DR Pfam: PF00078; RVT; 1.  
 DR RNA-directed DNA polymerase.  
 SQ SEQUENCE 608 AA; 70317 MW; BE89A7F0ED84AE73 CRC64;

Query Match 46.2%; Score 48.5; DB 5; Length 608;  
 Best Local Similarity 47.6%; Pred. No. 63;

Matches. 10; Conservative 5; Mismatches 5; Indels 1; Gaps 1;  
 Qy 1 KWKLKKPPPLKLLKL 20  
 Db 512 QWKQCLAFPPILLPILILEKM 532

## RESULT 10

ID 023898 PRELIMINARY; PRT; 608 AA.  
 AC 023898;  
 DT 01-NOV-1996 (TREMblrel. 01, Created)  
 DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)  
 DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)  
 DE Hypothetical 70.4 kDa protein.  
 OS Dictyostelium discoideum (Slime mold).  
 OC Eukaryota; Mycetozoa; Dictyostellida; Dictyostelium.  
 OX NCBI\_Taxid=44689;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE=86079481; PubMed=2416457;  
 RA Cappello J., Handelsman K., Lodish H.F.;  
 RT "Sequence of Dictyostelium DIRS-1: an apparent retrotransposon with  
 RT inverted terminal repeats and an internal circle junction sequence.";  
 RL Cell 43:105-115(1985).  
 DR EMBL: M11340; AAA70202.1; -.  
 DR InterPro: IPR000477; RVTse.  
 DR Pfam: PF00078; RVT; 1.  
 KW Hypothetical protein; RNA-directed DNA polymerase.  
 SQ SEQUENCE 608 AA; 70376 MW; B7285DAB3FE2FBF3 CRC64;

Query Match 46.2%; Score 48.5; DB 5; Length 608;  
 Best Local Similarity 47.6%; Pred. No. 63;  
 Matches 10; Conservative 5; Mismatches 5; Indels 1; Gaps 1;

Qy 1 KWKLKKPPPLKLLKL 20  
 Db 512 QWKQCLAFPPILLPILILEKM 532

## RESULT 11

ID 09HB1 PRELIMINARY; PRT; 700 AA.  
 AC 09HB1;  
 DT 01-MAR-2001 (TREMblrel. 16, Created)  
 DT 01-MAR-2001 (TREMblrel. 16, Last sequence update)  
 DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)  
 DE Calpain large polypeptide L2.  
 OS Homo sapiens (human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_Taxid=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=BRAIN ASTROCYTOMA;  
 RX MEDLINE=20403900; PubMed=10944468;  
 RA Ye Z., Connor J.R.;  
 RT "cDNA cloning by amplification of circularized first strand cDNAs  
 RT reveals non-IRE-regulated iron-responsive mRNAs.";  
 RL Biochem. Biophys. Res. Commun. 275:223-227(2000).  
 DR EMBL: AF261089; AAF99682.1; -.  
 DR HSP; P04574; IALV.  
 DR InterPro: IPR002048; EF-hand.  
 DR InterPro: IPR001300; Protease\_C2.  
 DR InterPro: IPR000169; SHprot\_acsite.  
 DR Pfam: PF01067; Calpain\_IIT; 1.  
 DR Pfam: PF00036; efhand; 3.  
 DR Pfam: PF00648; Peptidase\_C2; 1.  
 DR PRINTS: PR00704; CALPAIN.  
 DR SMART: SM00230; Cyspc; 1.  
 DR SMART: SM00054; Efh; 3.  
 DR PROSITE: PS00018; EF\_HAND; UNKNOWN\_1.  
 DR PROSITE: PS00139; THIOL\_PROTEASE\_CYS; 1.



SO SEQUENCE 700 AA; 79934 MW; DD304C603DEAC04C CRC64;  
 Query Match 46.2%; Score 48.5; DB 4; Length 700;  
 Best Local Similarity 52.6%; Pred. No. 72;  
 Matches 10; Conservative 5; Mismatches 3; Indels 1; Gaps 1;  
 QY 1 KKKKLLKKPPPLKKLLKK 19  
 DB 213 EWE-LKKPPNLFKIIQK 230  
 RESULT 12  
 ID 08WU26 PRELIMINARY; PRT; 700 AA.  
 AC 08WU26;  
 DT 01-MAR-2002 (TREMBLrel. 20, Created)  
 DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE Hypothetical 80.0 kDa protein.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-PANCREAS;  
 RA Strausberg R.;  
 RL Submitted (JAN-2002) to the EMBL/Genbank/DBJ databases.  
 DR EMBL; BC021303; AAH21303.1;  
 DR InterPro; IPR002048; EF-hand.  
 DR InterPro; IPR001300; Protease\_C2.  
 DR InterPro; IPR000169; SHprot\_acsite.  
 DR Pfam; PF01067; Calpain\_III; 1.  
 DR Pfam; PF00036; ehand; 3.  
 DR Pfam; PF00648; Peptidase\_C2; 1.  
 DR PRINTS; PR00704; CALPAIN.  
 DR SMART; SM00230; Cyspc; 1.  
 DR PROSITE; PS00018; EF\_HAND; UNKNOWN.1.  
 DR PROSITE; PS00139; THIOL\_PROTEASE\_CYS; UNKNOWN.1.  
 KW Hypothetical protein  
 SQ SEQUENCE 700 AA; 80009 MW; 8D105821A4374FB2 CRC64;  
 Query Match 46.2%; Score 48.5; DB 4; Length 700;  
 Best Local Similarity 52.6%; Pred. No. 72;  
 Matches 10; Conservative 5; Mismatches 3; Indels 1; Gaps 1;  
 QY 1 KKKKLLKKPPPLKKLLKK 19  
 DB 213 EWE-LKKPPNLFKIIQK 230  
 RESULT 13  
 ID 09GIG1 PRELIMINARY; PRT; 700 AA.  
 AC 09GIG1;  
 DT 01-MAR-2001 (TREMBLrel. 16, Created)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE Calpain 2.  
 OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;  
 OC Cercopithecinae; Macaca.  
 OX NCBI\_TaxID=9541;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-RETINA;  
 RA Nakajima T., Fukutake C., Azuma M., Shearer T.R.;  
 RT "Calpain isoforms in the eye of monkey."  
 RT Submitted (JUL-2000) to the EMBL/Genbank/DBJ databases.  
 DR EMBL; AF284441; AAG22771.1;  
 DR HSP; P04574; IALV.  
 DR MEROPS; C02.002; -

DR InterPro; IPR002048; EF-hand.  
 DR InterPro; IPR001300; Protease\_C2.  
 DR InterPro; IPR000169; SHprot\_acsite.  
 DR Pfam; PF01067; Calpain\_III; 1.  
 DR Pfam; PF00036; ehand; 3.  
 DR Pfam; PF00648; Peptidase\_C2; 1.  
 DR PRINTS; PR00704; CALPAIN.  
 DR SMART; SM00230; Cyspc; 1.  
 DR SMART; SM00054; Efn; 3.  
 DR PROSITE; PS00018; EF\_HAND; UNKNOWN.1.  
 DR PROSITE; PS00139; THIOL\_PROTEASE\_CYS; 1.  
 SQ SEQUENCE 700 AA; 80026 MW; DCEEL6214F05057C CRC64;  
 Query Match 46.2%; Score 48.5; DB 6; Length 700;  
 Best Local Similarity 52.6%; Pred. No. 72;  
 Matches 10; Conservative 5; Mismatches 3; Indels 1; Gaps 1;  
 QY 1 KKKKLLKKPPPLKKLLKK 19  
 DB 213 EWE-LKKPPNLFKIIQK 230  
 RESULT 14  
 ID 09YGC1 PRELIMINARY; PRT; 552 AA.  
 AC 09YGC1;  
 DT 01-MAY-1999 (TREMBLrel. 10, Created)  
 DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)  
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)  
 DE B cell linker protein BLNK.  
 GN BLNK OR BASH.  
 OS Gallus gallus (Chicken).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
 OC Gallus.  
 OX NCBI\_TaxID=9031;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE-99146381; PubMed-10023776;  
 RA Ishai M., Kurosaki M., Pappu R., Okawa K., Ronko I., Fu C.,  
 RA Shibata M., Yamatsu A., Chan A.C., Kurosaki T.;  
 RT "BLNK required for coupling Syk to PLC gamma 2 and Rac1-JNK in B cells."  
 RT Immunity 10:117-125(1999).  
 RL [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE-99049801; PubMed-9834055;  
 RA Gotsuka R., Fujimura Y., Mamada H., Umeda A., Morimura T.,  
 RA Uetsuka K., Doi K., Tsuji S., Kitamura D.;  
 RT "BASH, a novel signaling molecule preferentially expressed in B cells  
 of the bursa of Fabricius."  
 RT J. Immunol. 161:5804-5808(1998).  
 DR EMBL; AF089727; AAD12783.1;  
 DR EMBL; AB015289; BAA36275.1; -  
 DR HSP; P08487; 2PBD.  
 DR InterPro; IPR002865; P-rich\_extensn.  
 DR InterPro; IPR000980; SH2.  
 DR Pfam; PF00017; SH2; 1.  
 DR PRINTS; PR01217; PRICHEXTENS.  
 DR PRINTS; PR00401; SH2DOMAIN.  
 DR Prodom; PD000093; SH2; 1.  
 DR SMART; SM00252; SH2; 1.  
 DR PROSITE; PS00001; SH2; 1.  
 SQ SEQUENCE 552 AA; 61822 MW; FB232179BE38D072 CRC64;  
 Query Match 45.7%; Score 48; DB 13; Length 552;  
 Best Local Similarity 57.1%; Pred. No. 68;  
 Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;  
 QY 2 WKLLKKPPPLKK 15  
 DB 55 WDRLLKKPPPLPR 68

## RESULT 15

09PKS5 PRELIMINARY; PRT; 867 AA.  
 ID Q9PKS5;  
 AC Q9PKS5;  
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
 DE ATP-dependent Clp protease, subunit B.  
 GN TC0389.  
 OS Chlamydia muridarum.  
 OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.  
 OX NCBI\_Taxid=83560;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MOPN / NTG;  
 RX MEDLINE=20150255; Pubmed=10684935;  
 RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,  
 White O., Hickey E.K., Peterson J., Uterback T., Berry K., Bass S.,  
 Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,  
 Gwin M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,  
 Eisen J., Fraser C.M.;  
 RT "Genome sequences of Chlamydia trachomatis MOPN and Chlamydia  
 pneumoniae AR39.";  
 RL Nucleic Acids Res. 28:1397-1406(2000).  
 DR EMBL; AE002306; AAF39246.1; -.  
 DR TIGR; TC0389; -.  
 DR InterPro; IPR003593; AAA\_ATPase.  
 DR InterPro; IPR003959; AAA\_ATPase\_cent.  
 DR InterPro; IPR001270; Chaprinin\_clpA/B.  
 DR InterPro; IPR004176; Clp\_N.  
 DR Pfam; PF00004; AAA; 1.  
 DR Pfam; PF02861; Clp\_N; 2.  
 DR PRINTS; PR00300; CLPPROTEASEA.  
 DR SMART; SM00382; AAA; 1.  
 DR PROSITE; PS00870; CLPAB\_1; 1.  
 KW Complete proteome.  
 SQ SEQUENCE 867 AA; 96603 MW; DAB4C579AAC79F57 CRC64;

## Query Match

Best Local Similarity 45.7%; Score 48; DB 16; Length 867;

Matches 9; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

OY 2 WKLLKKPPPLKKLKL 20

DB 124 WKRTVKTSPAEKELIKL 142

Search completed: June 6, 2003, 13:24:54  
 Job time : 24.5 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 6, 2003, 13:21:00 ; Search time 7 Seconds

(without alignments)  
118.504 Million cell updates/sec

Title: US-10-081-418-2

Perfect score: 105

Sequence: 1 RKKRLKKPPPLKRLKLL 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 segs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database: SwissProt\_40:\*

Database: Lasting first 45 summaries

Database: Lasting first 45 summaries

Database: Lasting first 45 summaries

Database: Lasting first 45 summaries

Database: Lasting first 45 summaries

Database: Lasting first 45 summaries

Database: Lasting first 45 summaries

Database: Lasting first 45 summaries

Database: Lasting first 45 summaries

Database: Lasting first 45 summaries

Database: Lasting first 45 summaries

Database: Lasting first 45 summaries

Database: Lasting first 45 summaries

Database: Lasting first 45 summaries

Database: Lasting first 45 summaries

Database: Lasting first 45 summaries

Database: Lasting first 45 summaries

Database: Lasting first 45 summaries

Database: Lasting first 45 summaries

Database: Lasting first 45 summaries

# SUMMARIES

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result No.	Score	Query Match	Length	ID	Description
1	50	47.6	900	1 IF38_ARATH	049160 arabidopsis
2	49	46.7	110	1 Y107_NPVAC	P41660 autographa
3	49	46.7	1235	1 KPR2_RABIT	P46018 oryctolagus
4	48.5	46.2	700	1 CAN2_HUMAN	P17655 homo sapien
5	48	45.7	292	1 V226_FOWPV	Q91509 fowlpox vir
6	48	45.7	339	1 DCUP_HELPJ	Q92108 helicobacte
7	48	45.7	339	1 DCUP_HELPJ	O25325 helicobacte
8	47	44.8	1324	1 K167_HUMAN	P46013 homo sapien
9	46	43.8	366	1 M14A_DROME	P13026 hungarian g
10	46	43.8	537	1 GAG_BAEVM	O62618 drosophila
11	45.5	43.3	479	1 PBP4_HAEIN	P03341 baboon endo
12	45.5	43.3	700	1 CAN2_MOUSE	P45161 h penicilli
13	45.5	43.3	700	1 CAN2_MOUSE	O08529 mus musculi
14	45.5	43.3	700	1 CAN2_MOUSE	O07009 rattus norv
15	45	42.9	151	1 YP16_METTM	P14935 methanobact
16	45	42.9	1148	1 YK00_CAEEL	P34305 caenorhabdi
17	44.5	42.4	629	1 STR_PYRHO	O59147 pyrococcus
18	44	41.9	143	1 N1FX_FRAL	P46038 frankia aln
19	44	41.9	354	1 DCUP_PASMU	P57964 pasteurella
20	44	41.9	591	1 PUR9_HUMAN	P31939 homo sapien
21	44	41.9	779	1 CCA_YEAST	P07822 saccharomyc
22	44	41.9	1310	1 ACB_RABIT	P12822 oryctolagus
23	43	41.0	244	1 Y293_MYCGB	P47535 mycoplasma
24	43	41.0	487	1 MDM2_CANFA	P56950 canis famli
25	43	41.0	489	1 MDM2_MOUSE	P56951 mus musculi
26	43	41.0	491	1 MDM2_HORSE	O00987 homo sapien
27	43	41.0	491	1 MDM2_HUMAN	O94V48 vldrio choi
28	43	41.0	613	1 UR1D_VIBCH	O92178 gallus gall
29	43	41.0	700	1 CAN2_CHICK	P39964 schizosacch
30	43	41.0	757	1 CC5_SCHPO	P21190 saccharomyc
31	42.5	40.5	360	1 IME1_YEAST	P02777 bos taurus
32	42	40.0	88	1 PLE4_BOVIN	O14933 homo sapien
33	42	40.0	152	1 UBC8_HUMAN	

## ALIGNMENTS

34	42	40.0	173	1 YN5_CAEEL	P30644 caenorhabdi
35	42	40.0	282	1 TAUD_ECOLI	P37510 escherichia
36	42	40.0	343	1 B10B_BUCAL	P57378 buchnera ap
37	42	40.0	481	1 DD6_XENLA	P54824 xenopus lae
38	42	40.0	483	1 DD6_HUMAN	P26196 homo sapien
39	42	40.0	483	1 DD6_MOUSE	P54823 mus musculi
40	42	40.0	494	1 CP51_HUMAN	P08686 homo sapien
41	42	40.0	564	1 TP6B_PYRAB	O94135 pyrococcus
42	42	40.0	612	1 G1DA_MYCGB	P47619 mycoplasma
43	42	40.0	1068	1 YCFO_MARPO	P12221 marchantia
44	42	40.0	1192	1 K2_DROME	O46072 drosophila
45	42	40.0	1199	1 Y173_HUMAN	O14679 homo sapien

RESULT 1  
ID IF38\_ARATH STANDARD: PRT: 900 AA.  
AC 049160:

DT 30-MAY-2000 (rel. 39, Created)  
DT 30-MAY-2000 (rel. 39, Last sequence update)  
DT 16-OCT-2001 (rel. 40, Last annotation update)  
DE Probable eukaryotic translation initiation factor 3 subunit 8 (eif3 p110) (p110).  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=99065153; PubMed=9849901;  
RA Karniol B., Ianalom A., Kwok S., Tsuge T., Matsui M., Deng X.-W.,  
RA Chamovitz D.A.;  
RT "The Arabidopsis homologue of an eif3 complex subunit associates with  
the COP9 complex.";  
RT FEBS Lett. 439:173-179(1998).  
RL - FUNCTION: BINDS TO THE 40S RIBOSOME AND PROMOTES THE BINDING OF  
METHIONYL-TRNAI AND MRNA (BY SIMILARITY).  
CC - SUBUNIT: EIF-3 IS COMPOSED OF AT LEAST 10 DIFFERENT SUBUNITS (BY  
SIMILARITY).  
CC - SIMILARITY: BELONGS TO THE EIF38 FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sb.ch/announce/>  
or send an email to [license@isb-sb.ch](mailto:license@isb-sb.ch))  
CC  
CC EMBL: AF040102; AAC83464.1; -  
DR InterPro: IPR000717; PCT.  
DR Pfam: PF01399; PCT; 1.  
DR SMART: SM00088; PINT; 1.  
KW Initiation factor; Protein biosynthesis.  
SQ SEQUENCE 900 AA: 102989 MW: 36035PDB59AB50D CRC64;

Query Match Best Local Similarity 50.0%; Pred. No. 10;  
Matches 9; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

OY 2 WKRLKKPPPLKRLKLL 19  
DB 215 WEKMLSKDKLKLKMLNK 232

RESULT 2  
ID Y107\_NPVAC STANDARD: PRT: 110 AA.  
AC P41660;

```

01-NOV-1995 (Rel. 32, Created)
01-NOV-1995 (Rel. 32, Last sequence update)
01-NOV-1997 (Rel. 35, Last annotation update)
Hypothetical 12.5 kDa protein in HE65-PK2 intergenic region.
Autographa californica nuclear polyhedrosis virus (ACNPV).
Virus(es): dsDNA viruses, no RNA stage; Baculoviridae;
Nucleopolyhedrovirus.
NCBI_TaxID=46015;
[1]
SEQUENCE FROM N.A.
STRAIN=C6;
MEDLINE=94303173; PubMed=8030224;
Ayres M.D., Howard S.C., Kuzio J., Lopez-Ferber M., Possee R.D.;
"The complete DNA sequence of Autographa californica nuclear
polyhedrosis virus.";
Virology 202:586-605(1994).
-1- SIMILARITY: NO CORRESPONDING ORF IN OPMNPV, ORF106 AND ORF107 OF
ACNPV ARE FOUND AS A SINGLE ORF (ORF107) IN OPMNPV.
CC CC THE SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC EMBL: L22858; AAA6737.1; -.
DR Hypothetical protein.
KM
SQ SEQUENCE 110 AA; 12547 MW; 27B17C9DB21204FB CRC64;

Query Match 46.7%; Score 49; DB 1; Length 110;
Best Local Similarity 64.7%; Pred. No. 1.7;
Matches 11; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

0Y 3 KKLKKPPPLKLLKK 19
Db 85 KCLLPAPPPLKLEKK 101

RESULT 3
KPB2_RABIT
ID KPB2_RABIT STANDARD; PRT; 1235 AA.
AC P46018;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Phosphorylase B kinase alpha regulatory chain, liver isoform
DE (Phosphorylase kinase alpha L subunit).
GN PKRA2.
OS Oryctolagus cuniculus (Rabbit)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OC NCBI_TaxID=9986;
OX [1]
RN SEQUENCE FROM N.A.
RP TISSUE=Liver;
RC MEDLINE=92196064; PubMed=1372435;
RX Davidson J.J., Oezelik T., Hamacher C., Willems P.J., Francke U.,
RA Kilmann M.W.;
RA "CDNA cloning of a liver isoform of the phosphorylase kinase alpha
RT subunit and mapping of the gene to Xp22.2-p22.1, the region of human
RT X-linked liver glycogenosis.";
RT Proc. Natl. Acad. Sci. U.S.A. 89:2096-2100(1992).
CC -1- FUNCTION: PHOSPHORYLASE B KINASE CATALYZES THE PHOSPHORYLATION OF
CC SERINE IN CERTAIN SUBSTRATES, INCLUDING TROPONIN I. THE ALPHA
CC CHAIN MAY BIND CALMODULIN.
CC -1- ENZYME REGULATION: BY PHOSPHORYLATION OF VARIOUS SERINE RESIDUES
CC AND BY CALCIUM.
CC -1- PATHWAY: glycogen metabolism.
CC -1- SUBUNIT: POLYMER OF 16 CHAINS. FOUR EACH OF ALPHA, BETA, GAMMA,
CC AND DELTA. ALPHA AND BETA ARE REGULATORY CHAINS, GAMMA IS THE
CC CATALYTIC CHAIN, AND DELTA IS CALMODULIN.

```

```

CC -1- TISSU SPECIFICITY: PREDOMINANTLY EXPRESSED IN LIVER AND OTHER
CC NONMUSCLE TISSUES.
CC -1- SIMILARITY: BELONGS TO THE PHOSPHORYLASE B KINASE REGULATORY
CC CHAINS FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb.sib.ch).
CC -----
DR EMBL; X60421; CAA42952.1; -
RW Glycogen metabolism; Phosphorylation; Calmodulin-binding;
KW Multigene family.
FT DOMAIN 808 838 CALMODULIN-BINDING (POTENTIAL).
FT DOMAIN 1059 1099 CALMODULIN-BINDING (POTENTIAL).
SQ SEQUENCE 1235 AA; 138769 MW; B8E218C1D8C6F376 CRC64;
OY Query Match 46.7%; Score 49; DB 1; Length 1235;
Db Best Local Similarity 56.2%; Pred. No. 20;
Matches 9; Conservative 4; Mismatches 3; Indels 0; Gaps 0;
OY 5 LKKPPPLKKLKKL 20
ID 1; | | | | | : | | : |
CAN2_HUMAN STANDARD: PRT; 700 AA.
AC P17655; Q16738;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Calpain 2, large [catalytic] subunit precursor (EC 3.4.22.17)
DE (Calcium-activated neutral proteinase) (CAMP) (M-type) (M-calpain)
DE (Millimolar-calpain).
GN CAPN2 OR CAPNP2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89166474; PubMed=2852952;
RA Imafo S., Aoki K., Ohno S., Emori Y., Kawasaki H., Sugihara H.,
RA Suzuki K.;
RT "Molecular cloning of the cDNA for the large subunit of the
RT high-Ca2+-requiring form of human Ca2+-activated neutral protease.";
RL Biochemistry 27:8122-8128(1988).
RN [2]
RP SEQUENCE OF 1-79 FROM N.A.
RC TISSUE=Lymph node;
RX MEDLINE=88197947; PubMed=2539381;
RA Hata A., Ohno S., Akita Y., Suzuki K.;
RT "Randomly reterized negative enhancer-like elements regulate
RT transcription of a human gene for the large subunit of calcium-
RT dependent protease.";
RL J. Biol. Chem. 264:6404-6411(1989).
RN [3]
RN X-RAY CRYSTALLOGRAPHY (2.7 ANGSTROMS).
RX MEDLINE=20105516; PubMed=10639123;
RA Strobl S., Fernandez-Catalan C., Braun M., Huber R., Masumoto H.,
RA Nakagawa K., Irie A., Sorimachi H., Bourenkow G., Bartunik H.,
RA Suzuki K., Bode W.;
RT "The crystal structure of calcium-free human m-calpain suggests an
RT electrostatic switch mechanism for activation by calcium.";
RL Proc. Natl. Acad. Sci. U.S.A. 97:588-592(2000).
CC -1- FUNCTION: Calcium-regulated non-Lysosomal thiol-protease which
CC catalyze limited proteolysis of substrates involved in
CC cytoskeletal remodelling and signal transduction.

```

CC -1- CATALYTIC ACTIVITY: Preferential cleavage: Tyr-|-Xaa, Met-|-Xaa or  
 CC Arg-|-Xaa with Leu or Val as the p2 residue.  
 CC -1- COFACTOR: Binds 3 calcium ions.  
 CC -1- ENZYME REGULATION: Activated by 200-1000 micromolar concentrations  
 CC of calcium and inhibited by calpastatin.  
 CC -1- SUBUNIT: Heterodimer of a large (catalytic) and a small  
 CC (regulatory) subunit.  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic; Translocates to the plasma  
 CC membrane upon Ca++ binding.  
 CC -1- SIMILARITY: CONTAINS 5 EF-HAND CALCIUM-BINDING DOMAINS.  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C2.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-slb.ch/announce/>  
 CC or send an email to [license@isb-slb.ch](mailto:license@isb-slb.ch)).  
 CC -----  
 CC EMBL; M23234; AAA35645.1; -  
 CC EMBL; J04700; AAA52760.1; -  
 CC PIR; A31218; A31218.  
 CC PDB; 1KRU; 07-DEC-01.  
 CC MEROPS; C02.002; -  
 CC Genew; HGNC:1479; CAPN2.  
 CC MIM; 114230; -  
 CC InterPro; IPR002048; EF-hand.  
 CC InterPro; IPR001300; Protease\_C2.  
 CC InterPro; IPR00169; SHprot\_acsite.  
 CC Pfam; PF00036; efhand; 3.  
 CC Pfam; PF00648; peptidase\_C2; 1.  
 CC Pfam; PF01067; Calpain\_III; 1.  
 CC PRINTS; PR00704; CALPAIN.  
 CC ProDom; PD000012; EF-hand; 1.  
 CC SMART; SM00230; Cyspc; 1.  
 CC SMART; SM00054; Efh; 2.  
 CC PROSITE; PS00018; EF-HAND; 2.  
 CC PROSITE; PS00139; THIOL\_PROTEASE\_CYS; 1.  
 CC PROSITE; PS00639; THIOL\_PROTEASE\_HIS; FALSE NEG.  
 CC PROSITE; PS00640; THIOL\_PROTEASE\_ASN; FALSE NEG.  
 CC K1 Hydroxylase: Thiol protease; Calcium-binding; Repeat; Multigene family;  
 CC 3D-structure.  
 CC FT PROPEP 1 19 ANCHORS TO THE SMALL SUBUNIT (POTENTIAL).  
 CC FT CHAIN 20 700 CALPAIN 2, LARGE (CATALYTIC) SUBUNIT.  
 CC FT DOMAIN 20 210 THIOL\_PROTEASE DOMAIN I.  
 CC FT DOMAIN 211 355 THIOL\_PROTEASE DOMAIN II.  
 CC FT DOMAIN 356 514 THIOL\_PROTEASE DOMAIN III.  
 CC FT DOMAIN 515 529 THIOL\_PROTEASE DOMAIN IV.  
 CC FT DOMAIN 530 700 LINKER.  
 CC FT CA\_BIND 541 552 DOMAIN IV.  
 CC FT CA\_BIND 552 552 EF-HAND 1.  
 CC FT CA\_BIND 585 596 EF-HAND 2.  
 CC FT CA\_BIND 615 626 EF-HAND 3.  
 CC FT DOMAIN 650 661 ANCESTRAL CALCIUM SITE 3 (POTENTIAL).  
 CC FT DOMAIN 680 691 ANCESTRAL CALCIUM SITE 4 (POTENTIAL).  
 CC FT ACT\_SITE 105 105 BY SIMILARITY.  
 CC FT ACT\_SITE 262 262 BY SIMILARITY.  
 CC FT ACT\_SITE 286 286 BY SIMILARITY.  
 CC FT CONFLICT 68 68 S -> G (IN REF. 2).  
 CC FT CONFLICT 73 74 MR -> IE (IN REF. 2).  
 CC SO SEQUENCE 700 AA; 80006 MW; A944D13BC8465531 CRC64;

OY 1 KKKKKKKPPLKKLKK 19  
 DB 213 EWE-LKKPPNLEKTIKK 230

Query Match 46.2%; Score 48.5; DB 1; Length 700;  
 Best Local Similarity 52.6%; Pred. No. 13;  
 Matches 10; Conservative 5; Mismatches 3; Indels 1; Gaps 1;

ID V226\_F0MPV STANDARD; PRT; 292 AA.  
 AC 09J509;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Probable serine/threonine protein kinase FPV226 (EC 2.7.1.-).  
 GN FPV226.  
 OS Fowlpox virus (FPV).  
 OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;  
 OC Avipoxvirus.  
 OX NCBI\_TaxID=10261;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20193820; PubMed=10729156;  
 RA Afonso C.L., Tjulan E.R., Lu Z., Zsak L., Kutish G.F., Rock D.L.;  
 RT "The genome of fowlpox virus."  
 RL J. Virol. 74:3815-3831(2000).  
 CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.  
 CC POXYVIRUSES SUBFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-slb.ch/announce/>  
 CC or send an email to [license@isb-slb.ch](mailto:license@isb-slb.ch)).  
 CC -----  
 CC EMBL; AF196100; AAF4570.1; -  
 CC InterPro; IPR000719; Euk\_pkinase.  
 CC InterPro; IPR002290; Ser\_thr\_pkinase.  
 CC Pfam; PF00069; pkinase; 1.  
 CC ProDom; PD000001; Euk\_pkinase; 1.  
 CC PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
 CC PROSITE; PS00108; PROTEIN\_KINASE\_ST; 1.  
 CC PROSITE; PS00011; PROTEIN\_KINASE\_DOM; 1.  
 CC TRANSFERASE: Serine/threonine-protein kinase; ATP-binding.  
 CC FT DOMAIN 14 292 PROTEIN KINASE.  
 CC FT NP\_BIND 20 28 ATP (BY SIMILARITY).  
 CC FT BINDING 43 43 ATP (BY SIMILARITY).  
 CC FT ACT\_SITE 147 147 BY SIMILARITY.  
 CC SO SEQUENCE 292 AA; 33990 MW; CC0C0A9632D3BD CRC64;

OY 4 KKKKKPPLKKLKKL 20  
 DB 115 KKKKKPPLKKLKKL 131

Query Match 45.7%; Score 48; DB 1; Length 292;  
 Best Local Similarity 47.1%; Pred. No. 6.5;  
 Matches 8; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

RESULT 6  
 DCOP\_HELPJ STANDARD; PRT; 339 AA.  
 ID DCOP\_HELPJ  
 AC 09ZLM8;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Uroporphyrinogen decarboxylase (EC 4.1.1.37) (URO-D) (UPD).  
 GN HEMP OR JHP0551.  
 OS Helicobacter pylori J99 (Campylobacter pylori J99).  
 OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;  
 OC Helicobacter.  
 OX NCBI\_TaxID=85963;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=99120557; PubMed=9923682;  
 RA Alm R.A., Ling L.-S.L., Moir D.T., Dolg P.C.,  
 RA Smith D.R., Noonan B., Guild B.C., deJonghe B.L., Carmel G.,  
 RA Tummino P.J., Caruso A., Uria-Nickelsen M., Mills D.M., Ives C.,  
 RA Gibson R., Merberg D., Mills S.D., Jiang Q., Taylor D.E., Voyis G.F.,  
 RA Trust T.J.;

```

RT "Genomic sequence comparison of two unrelated isolates of the human
RT gastric pathogen Helicobacter pylori."
RL Nature 397:176-180(1999).
CC -1- CATALYTIC ACTIVITY: Uroporphyrinogen-III - coproporphyrinogen + 4
CC CO(2).
CC -1- PATHWAY: Porphyrin biosynthesis.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC -1- SIMILARITY: BELONGS TO THE UROPORPHYRINOGEN DECARBOXYLASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE001487; AAD06123.1; -.
DR HSSP; P06132; URO.
DR InterPro: IPR000257; Uro-decarboxyls.
DR Pfam; PF01208; URO-D; 1.
DR PROSITE; PS00906; UROD_1; 1.
DR PROSITE; PS00907; UROD_2; 1.
DR Lyase; Decarboxylase; Porphyrin biosynthesis; Complete proteome.
SQ SEQUENCE 339 AA; 38367 MW; 2BDDFF4D46C54 CRC64;

Query Match 45.7%; Score 48; DB 1; Length 339;
Best Local Similarity 55.0%; Pred. No. 7.6;
Matches 11; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 1 KWKLLKKPPPLKLLKLL 20
   11111 11111111
Db 159 KSKKMLYSEPEVKALLKLL 178

RESULT 7
DCUP_HELPY STANDARD; PRT; 339 AA.
AC 025325;
DT 15-JUN-1998 (Rel. 36, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Uroporphyrinogen decarboxylase (EC 4.1.1.37) (URO-D) (UPD).
GN HEME OR HP0604.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
CC Helicobacter.
CX NCBI_TaxID=210;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=26695 / ATCC 700392;
RX MEDLINE=97394467; Pubmed=9252185;
RA Tomb J.-F., White O., Kerlavage A.R., Clayton R.A., Sutton G.G.,
RA Fleischmann R.D., Ketchum K.A., Klenk H.-P., Gill S., Dougherty B.A.,
RA Nelson K., Quackenbush J., Zhou L., Kirness E.F., Peterson S.,
RA Loftus B., Richardson D., Dodson R., Khalak H.G., Glodek A.,
RA McInerney K., Fitzgerald L.M., Lee N., Adams M.D., Hickey E.K.,
RA Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,
RA Cotton M.D., Meldrum J.M., Fujii C., Bowman C., Matthey L., Wallin E.,
RA Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,
RA Venter J.C.;
RT "The complete genome sequence of the gastric pathogen Helicobacter
RT pylori."
RL Nature 388:539-547(1997).
CC -1- CATALYTIC ACTIVITY: Uroporphyrinogen-III - coproporphyrinogen + 4
CC CO(2).
CC -1- PATHWAY: Porphyrin biosynthesis.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC -1- SIMILARITY: BELONGS TO THE UROPORPHYRINOGEN DECARBOXYLASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

```

```

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE000574; AAD07669.1; ALT_INIT.
DR HSSP; P06132; URO.
DR TIGR; HP0604; -.
DR InterPro: IPR000257; Uro-decarboxyls.
DR Pfam; PF01208; URO-D; 1.
DR PROSITE; PS00906; UROD_1; 1.
DR PROSITE; PS00907; UROD_2; 1.
DR Lyase; Decarboxylase; Porphyrin biosynthesis; Complete proteome.
SQ SEQUENCE 339 AA; 38339 MW; 76B03ADAA42726 CRC64;

Query Match 45.7%; Score 48; DB 1; Length 339;
Best Local Similarity 55.0%; Pred. No. 7.6;
Matches 11; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 1 KWKLLKKPPPLKLLKLL 20
   11111 11111111
Db 159 KSKKMLYSEPEVKALLKLL 178

RESULT 8
KI67_HUMAN STANDARD; PRT; 3256 AA.
AC P46013;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Antigen Ki-67.
GN MKI67.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
CX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94043435; Pubmed=8227122;
RA Schlieter C., Duchrow M., Wohlenberg C., Becker M.H.G., Key G.,
RA Flad H.-D., Gerdes J.;
RT "The cell proliferation-associated antigen of antibody Ki-67: a very
RT large, ubiquitous nuclear protein with numerous repeated elements,
RT representing a new kind of cell cycle-maintaining proteins."
RL J. Cell Biol. 123:513-522(1993).
RN [2]
RP SEQUENCE OF 1-31 FROM N.A.
RA Gerdes J.;
RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: THOUGHT TO BE REQUIRED FOR MAINTAINING CELL
CC PROLIFERATION.
CC -1- SUBCELLULAR LOCATION: NUCLEAR. PREDOMINANTLY LOCALIZED IN THE G1
CC PHASE IN THE PERINUCLEAR REGION, IN THE LATER PHASES IT IS ALSO
CC DETECTED THROUGHOUT THE NUCLEAR INTERIOR, BEING PREDOMINANTLY
CC LOCALIZED IN THE NUCLEAR MATRIX. IN MITOSIS, IT IS PRESENT ON ALL
CC CHROMOSOMES.
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS: A LONG FORM (SHOWN HERE) AND A
CC SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- DEVELOPMENTAL STAGE: EXPRESSION OF THIS ANTIGEN OCCURS
CC PREFERENTIALLY DURING LATE G1, S, G2, AND M PHASES OF THE CELL
CC CYCLE, WHILE IN CELLS IN G0 PHASE THE ANTIGEN CANNOT BE DETECTED.
CC -1- SIMILARITY: CONTAINS 1 FHA DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X65550; CAA46519.1; -.

```

DR EMBL: X65551; CAA46520.1; -  
 DR EMBL: X94762; CAA64388.1; -  
 DR Genem; HGNC:7107; MK167.  
 DR MIM: 167741; -  
 DR InterPro: IPR000253; FHA\_domain.  
 DR Pfam: PF00498; FHA; 1.  
 DR SMART; SM00240; FHA; 1.  
 DR PROSITE; PS50006; FHA\_DOMAIN; 1.  
 DR Cell cycle; Antigen; Nuclear protein; ATP-binding; Repeat;  
 DR Alternative splicing.  
 FT DOMAIN 27  
 FT 1000 2928 FHA.  
 FT REPEAT 1000 1112 16 X 122 AA APPROXIMATE REPEATS.  
 FT REPEAT 1000 1122 1.  
 FT REPEAT 1234 1234 2.  
 FT REPEAT 1244 1356 3.  
 FT REPEAT 1366 1477 4.  
 FT REPEAT 1487 1598 5.  
 FT REPEAT 1608 1720 6.  
 FT REPEAT 1730 1842 7.  
 FT REPEAT 1851 1964 8.  
 FT REPEAT 1974 2086 9.  
 FT REPEAT 2096 2204 10.  
 FT REPEAT 2214 2326 11.  
 FT REPEAT 2335 2447 12.  
 FT REPEAT 2457 2569 13.  
 FT REPEAT 2579 2689 14.  
 FT REPEAT 2699 2808 15.  
 FT REPEAT 2818 2928 16.  
 FT NP\_BIND 3034 3041 ATP (POTENTIAL).  
 FT VASPLC 136 495 MISSING (IN SHORT ISOFORM).  
 SQ SEQUENCE 3256 AA; 358741 MW; 578F8C51BED42517 CRC64;

Query Match 45.7%; Score 48; DB 1; Length 3256;  
 Best Local Similarity 47.4%; Pred. No. 74;  
 Matches 9; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

OY 1 KWKLKKPPPLKKLKK 19  
 DB 535 KRKSLVMTTPVLKKIKE 553

RESULT 9  
 POL2\_GCMV STANDARD; PRT; 1324 AA.  
 AC P13026;  
 DT 01-JAN-1990 (Rel. 13, Created)  
 DT 01-JAN-1990 (Rel. 13, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE RN2 polypeptide (147 kDa protein) [Contains: Coat protein; 84 kDa protein].  
 OS Hungarian grapevine chrome mosaic virus (GCMV).  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Comoviridae;  
 OC Nepovirus.  
 OC NCB1\_TaxID=12273;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 814-822.  
 RA MEDLINE-90016865; PubMed-2798129;  
 RA Brault V., Hibrand L., Candresse T., Le Gall O., Dunez J.;  
 RT "Nucleotide sequence and genetic organization of Hungarian grapevine chrome mosaic nepovirus RN2";  
 RL Nucleic Acids Res. 17:7809-7823(1989).  
 CC -1- FUNCTION: THE PROTEIN LOCATED AT THE N-TERMINAL END OF THE NEPOVIRUS POLYPEPTIDE COULD BE REQUIRED FOR NEMATODE TRANSMISSION OF THE VIRUS.  
 CC -1- PTM: THE RN2 POLYPEPTIDE IS CLEAVED BY AN RNASE-ENCODED PROTEASE TO YIELD THE MATURE COAT PROTEIN AND AN 84 KDA PROTEIN WHICH IS FURTHER CLEAVED INTO TWO PRODUCTS OF APPROXIMATELY 46 AND 48 KDA.  
 CC -1- SIMILARITY: TO THE RN2 POLYPEPTIDE OF OTHER NEPOVIRUSES.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: X15163; CAA33255.1; -  
 DR PIR: S06187; S06187.  
 DR InterPro: IPR005054; Nepo\_coat.  
 DR InterPro: IPR005305; Nepo\_coat\_C.  
 DR InterPro: IPR005306; Nepo\_coat\_N.  
 DR Pfam: PF03391; Nepo\_coat; 1.  
 DR Pfam: PF03688; Nepo\_coat\_C; 1.  
 DR Pfam: PF03689; Nepo\_coat\_N; 1.  
 DR Polypeptide; Coat protein.  
 FT MOD\_RES 811 811  
 FT DOMAIN 1 810 BLOCKED (PROBABLE).  
 FT 811 1324 84 KDA PROTEIN.  
 FT VARIANT 400 400 COAT PROTEIN.  
 FT VARIANT 1006 1006 N -> D.  
 FT VARIANT 1006 1006 I -> T.  
 SQ SEQUENCE 1324 AA; 147804 MW; 302C286B2EA1756A CRC64;

Query Match 44.8%; Score 47; DB 1; Length 1324;  
 Best Local Similarity 52.9%; Pred. No. 42;  
 Matches 9; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

OY 2 WKKLKKPPPLKKLKK 18  
 DB 1158 WFTLLRPDPKSLKILK 1174

RESULT 10  
 M14A\_DROME STANDARD; PRT; 366 AA.  
 ID M14A\_DROME  
 AC 062618; O62618; O9TYB4;  
 DT 15-JUN-2002 (Rel. 41, Created)  
 DT 15-JUN-2002 (Rel. 41, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Mitogen-activated protein kinase 14A (EC 2.7.1.-) (MAP kinase p38a)  
 DE (p38 MAPK) (Dp38) (D-p38a).  
 GN MPK3 OR p38A OR CG5475.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;  
 OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
 OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
 OC NCB1\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A., FUNCTION, DEVELOPMENTAL STAGE, AND ENZYME REGULATION.  
 RP TISSUE-Embryo;  
 RC MEDLINE-98252940; PubMed-9584193;  
 RA Han Z.S., Enslin H., Hu X., Meng X., Wu I.-H., Barrett T., Davis R.J.,  
 RA Ip Y.T.;  
 RT "A conserved p38 mitogen-activated protein kinase pathway regulates Drosophila immunity gene expression";  
 RL Mol. Cell. Biol. 18:3527-3539(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A., PHOSPHORYLATION OF TYR-186, FUNCTION, SUBCELLULAR LOCATION, AND ENZYME REGULATION.  
 RP TISSUE-Hemocyte;  
 RC MEDLINE-98073070; PubMed-9417090;  
 RA Han S.-J., Choi K.-Y., Brey P.T., Lee W.-J.;  
 RT "Molecular cloning and characterization of a Drosophila p38 mitogen-activated protein kinase";  
 RL J. Biol. Chem. 273:369-374(1998).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RP STRAIN-Berkeley;  
 RC MEDLINE-20196006; PubMed-10731132;  
 RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazey R.G., Champs M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,

RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borokva D., Boichain M.R., Bouck J., Brokstein P., Brothier P.,  
 RA Butts R.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahle K., Davenport L.B., Davies P.,  
 RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrelia S., Fleischmann W.,  
 RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Fiodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howard T.J., Wei M.-H., Idegawa C.,  
 RA Jalali M., Kalush F., Kapen G.H., Ke Z., Kenison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Mishina N.V., Mobarri C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M.D., Skupski M.P., Smith T.,  
 RA Slier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Sytkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Morley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RA "The genome sequence of Drosophila melanogaster";  
 RA Science 287:2185-2195(2000).  
 RA [4]  
 RN SEQUENCE OF 38-188 FROM N.A.  
 RP TISSUE-imaginal disks;  
 RC MEDLINE=9235284; PubMed=1378625;  
 RX Biggs W.H. III, Zilpursky S.L.;  
 RT "Primary structure, expression, and signal-dependent tyrosine  
 phosphorylation of a Drosophila homolog of extracellular signal-  
 regulated kinase.";  
 RT Proc. Natl. Acad. Sci. U.S.A. 89:6295-6299(1992).  
 RL -1- FUNCTION: Kinase involved in a signal transduction pathway. May  
 down regulate insect immunity gene expression after prolonged  
 infection.  
 CC -1- ENZYME REGULATION: Activated by MKK3 and by phosphorylation on  
 Tyr-186 in response to environmental stress.  
 CC -1- SUBCELLULAR LOCATION: Nuclear.  
 CC -1- DEVELOPMENTAL STAGE: Expressed both maternally and zygotically.  
 CC Levels are highest at the preblastoderm stage but low levels are  
 present throughout development.  
 CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.  
 CC MAP KINASE SUBFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL Outstation  
 at the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: AF035546; AAC39030.1; -;  
 DR EMBL: AF035547; AAC39031.1; -;  
 DR EMBL: U68667; AAB97138.1; -;  
 DR EMBL: AE003746; AAF56244.1; -;  
 DR HSP: Q16539; IWC.  
 DR FLYBase: FBgn0015765; MPK2.  
 DR InterPro: IPR000719; MAP\_kinase.  
 DR InterPro: IPR002290; Ser\_thr\_kinase.  
 DR Pfam: PF000069; pkinase.1.  
 DR ProDom: PD000001; Euk\_pkinase.1.  
 DR SMART: SM00220; S\_TKc.1.

DR PROSITE: PS01351; MAPK.1.  
 DR PROSITE: PS00107; PROTEIN\_KINASE\_ATP.1.  
 DR PROSITE: PS00011; PROTEIN\_KINASE\_DOM.1.  
 DR PROSITE: PS00108; PROTEIN\_KINASE\_ST; FALSE\_NEG.  
 KW Transferase; Serine/threonine-protein kinase; ATP-binding;  
 KW Nuclear protein; Phosphorylation.  
 FT DOMAIN 25 312  
 FT NP\_BIND 31 39  
 FT BINDING 54 54  
 FT ACT\_SITE 154 154  
 FT MOD\_RES 184 184  
 FT MOD\_RES 186 186  
 FT CONFLICT 27 27  
 FT CONFLICT 77 77  
 FT CONFLICT 80 80  
 FT CONFLICT 108 108  
 FT CONFLICT 124 124  
 FT CONFLICT 149 149  
 FT CONFLICT 163 163  
 SQ SEQUENCE 366 AA; 4225 MW; B3592B869F97990E CRC64;  
 Query Match 43.8%; Score 46; DB 1; Length 366;  
 Best Local Similarity 41.7%; Pred. No. 16;  
 Matches 10; Conservative 4; Mismatches 4; Indels 6; Gaps 1;  
 QY 1 KMKRLK-----KPPLLKRLK 18  
 DB 340 KMKELIKETVNFKPPSYAQVLK 363  
 ID GAG\_BAEVM STANDARD; PRT; 537 AA.  
 AC P03341; P10268;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 01-MAR-1989 (Rel. 10, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE GAG polypeptide [Contains: Inner coat protein P12; Core protein P15;  
 DE Core shell protein P30; Nucleoprotein P10].  
 GN GAG.  
 OS Baboon endogenous virus (strain M7).  
 OC Viruses; Retroviral viruses; Retroviridae; Mammalian type C retroviruses.  
 OX NCBI\_TaxID=11764;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Kato S., Matsuo K., Nishimura N., Takahashi N., Takano T.;  
 RT "The entire nucleotide sequence of baboon endogenous virus DNA: a  
 RT chimeric genome structure of murine type C and simian type D  
 RT retroviruses.";  
 RL Jpn. J. Genet. 62:127-137(1987).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=83241915; PubMed=6408267;  
 RA Tamura T.;  
 RT "Provirus of M7 baboon endogenous virus: nucleotide sequence of the  
 RT gag-pol region.";  
 RL J. Virol. 47:137-145(1983).  
 CC -1- PPM: SPECIFIC ENZYMATIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.  
 CC -1- MISCELLANEOUS: THIS PROTEIN IS SYNTHESIZED AS A GAG-POL  
 POLYPEPTIDE.  
 CC -1- SIMILARITY: CONTAINS 1 CCHC-TYPE ZINC FINGER  
 CC -1- CAUTION: REF. 2 SEQUENCE DIFFERS FROM THAT SHOWN EXTENSIVELY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL Outstation  
 at the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: D10032; BAA00923.1; -;  
 DR EMBL: X05470; CAA29027.1; -;



DR EMBL: J02034; -- NOT\_ANNOTATED\_CDS.  
 DR PIR: A03939; FOMVVB.  
 DR PIR: J0260; FOMVVB.  
 DR InterPro: IPR000840; Gag\_MA.  
 DR InterPro: IPR000306; Gag\_P30.  
 DR InterPro: IPR001878; Znf\_CCHC.  
 DR Pfam: PF00098; Znf\_CCHC; 1.  
 DR Pfam: PF01140; Gag\_MA; 1.  
 DR Pfam: PF02093; Gag\_P30; 1.  
 DR SMART: SM00343; Znf\_C2HC; 1.  
 DR PROSITE: PS50158; Zf\_CCHC; 1.  
 KW Core protein; Coat protein; Nucleoprotein; Polyprotein; Myristate;  
 ZINC-finger.  
 FT CHAIN 2 112 INNER COAT PROTEIN P12.  
 FT CHAIN 113 226 CORE PROTEIN P15.  
 FT CHAIN 227 477 CORE SHELL PROTEIN P30.  
 FT CHAIN 478 537 NUCLEOPROTEIN P10.  
 FT 2N-FING 501 518 CCHC-TYPE.  
 FT LIPID 2 MYRISTATE.  
 SQ SEQUENCE 537 AA; 60623 MW; 5DEEE4437CFCFB79 CRC64;  
 Query Match 43.8%; Score 46; DB 1; Length 537;  
 Best Local Similarity 43.8%; Pred. No. 23;  
 Matches 7; Conservative 4; Mismatches 5; Indels 0; Gaps 0;  
 OY 2 WKLLKKPPLKKLL 17  
 DB 86 MEDLVNPPVWVAPFL 101  
 ID PBP4\_HAEIN STANDARD; PRT; 479 AA.  
 AC P45161;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Penicillin-binding protein 4 precursor (PBP-4) [Includes: D-alanyl-D-  
 alanine carboxypeptidase (EC 3.4.16.4) (DD-peptidase) (DD-  
 carboxypeptidase); D-alanyl-D-alanine-endopeptidase (EC 3.4.99.-) (DD-  
 endopeptidase)].  
 GN DADB OR H11330.  
 OS Haemophilus influenzae.  
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
 OC Haemophilus.  
 OX NCBI\_Taxid=727;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-Rd / KM20 / ATCC 51907;  
 RX MEDLINE=95350630; PubMed=7542800;  
 RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,  
 RA Kesteven A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,  
 RA McKenny K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,  
 RA Scott J.D., Shirley R., Liu L.-T., Glodek A., Kelley J.M.,  
 RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,  
 RA Usterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,  
 RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geophagen N.S.M.,  
 RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,  
 RA Venter J.C.;  
 RT "Whole-genome random sequencing and assembly of Haemophilus influenzae  
 Rd";  
 RL Science 269:496-512(1995).  
 CC -1- FUNCTION: NOT INVOLVED IN TRANSPEPTIDATION BUT EXCLUSIVELY  
 CC CATALYZES A DD-CARBOXYPEPTIDASE AND DD-ENDOPEPTIDASE REACTION  
 CC (BY SIMILARITY).  
 CC -1- CATALYTIC ACTIVITY: D-alanyl-D-alanine + H(2)O -> 2 D-alanine.  
 CC -1- SUBSTRATE: FINAL STAGES IN PEPTIDOGLYCAN SYNTHESIS.  
 CC -1- PATHWAY: CELLULOSE DEGRADATION (POTENTIAL).  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S13; ALSO KNOWN AS THE  
 CC D-ALANYL-D-ALANINE CARBOXYPEPTIDASE 3 FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sdb.ch/announce/>  
 CC or send an email to [license@sdb-sdb.ch](mailto:license@sdb-sdb.ch)).  
 CC -----  
 DR EMBL: U32812; AAC22975.1; --  
 DR MEROPS: S13.001; --  
 DR TIGR: H11330; --  
 DR InterPro: IPR000667; Peptidase\_S13.  
 DR Pfam: PF02113; Peptidase\_S13; 1.  
 DR PRINTS: PR00922; DADACHPASE3.  
 DR TRIGRAMS: TRIGR00666; PBP4; 1.  
 KW Peptidoglycan synthetase; Cell division; Cell wall; Hydrolase; Signal;  
 KW Antibiotic resistance; Periplasmic; Complete proteome.  
 FT SIGNAL 1 26  
 FT CHAIN 27 479  
 FT ACT\_SITE 69 69  
 FT ACT\_SITE 69 69  
 FT ACT\_SITE 420 420  
 FT ACT\_SITE 420 420  
 SQ SEQUENCE 479 AA; 52685 MW; 632868C61206CB48 CRC64;  
 Query Match 43.3%; Score 45.5; DB 1; Length 479;  
 Best Local Similarity 64.7%; Pred. No. 25;  
 Matches 11; Conservative 2; Mismatches 3; Indels 1; Gaps 1;  
 OY 4 KLLKP-PPLKKLLK 19  
 DB 293 KLLSRPLDLKKMKK 309  
 ID CAN2\_MOUSE STANDARD; PRT; 700 AA.  
 AC O08529; O35518; O54843;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Calpain 2, large [catalytic] subunit precursor (EC 3.4.22.17)  
 DE (Calcium-activated neutral proteinase) (CANP) (M-type) (M-calpain)  
 DE (Molecular-calpain) (80 kDa M-calpain subunit) (CALP80).  
 GN CAPN2.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_Taxid=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BALB/c;  
 RX MEDLINE=97480729; PubMed=9339374;  
 RA Dear N., Matena K., Vingron M., Boehm T.;  
 RT "A new subfamily of vertebrate calpains lacking a calmodulin-like  
 RT domain: implications for calpain regulation and evolution";  
 RL Genomics 45:175-184(1997).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BALB/c;  
 RA Ozaki Y.;  
 RT submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.  
 RL [3]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=CNS;  
 RA Glass J.D., Nash N.R., Dry I., Culver D., Wesselingh S.;  
 RT "Cloning of m-calpain from mouse nervous system";  
 RL submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: Calcium-regulated non-lysosomal thiol-protease which  
 CC catalyze limited proteolysis of substrates involved in  
 CC cytoskeletal remodelling and signal transduction (by similarity).  
 CC -1- CATALYTIC ACTIVITY: Preferential cleavage: Tyr-I-Xaa, Met-I-Xaa or  
 CC Arg-I-Xaa with Leu or Val as the p2 residue.  
 CC -1- COFACTOR: Binds 3 calcium ions.  
 CC -1- ENZYME REGULATION: Activated by 200-1000 micromolar concentrations  
 CC of calcium and inhibited by calpastatin.



DR Pfam; PF00036; ehand; 3.  
 DR Pfam; PF00648; Peptidase\_C2; 1.  
 DR Pfam; PF01067; Calpain\_III; 1.  
 DR PRINTS; PR00704; CALPAIN.  
 DR Prodom; PD000012; EF-hand; 1.  
 DR SMART; SM00230; Cyspc; 1.  
 DR SMART; SM00054; Efh; 2.  
 DR PROSITE; PS00018; EF\_HAND; 2.  
 DR PROSITE; PS00139; THIOL\_PROTEASE\_CYS; 1.  
 DR PROSITE; PS00639; THIOL\_PROTEASE\_HIS; FALSE\_NEG.  
 DR PROSITE; PS00640; THIOL\_PROTEASE\_ASN; FALSE\_NEG.  
 KW Hydrolase; Thiol protease; Calcium-binding; Repeat; Multigene family;  
 KM 3D-structure.  
 FT PROPEP 1 19 ANCHORS TO THE SMALL SUBUNIT (POTENTIAL).  
 FT CHAIN 20 700 CALPAIN 2, LARGE (CATALYTIC) SUBUNIT.  
 FT DOMAIN 20 210 THIOL\_PROTEASE DOMAIN I.  
 FT DOMAIN 211 355 THIOL\_PROTEASE DOMAIN II.  
 FT DOMAIN 356 514 THIOL\_PROTEASE DOMAIN III.  
 FT DOMAIN 515 529 DOMAIN III, C2-LIKE DOMAIN.  
 FT DOMAIN 530 700 LINKER.  
 FT CA\_BIND 541 552 DOMAIN IV.  
 FT CA\_BIND 585 596 EF\_HAND 1.  
 FT CA\_BIND 615 626 EF\_HAND 2.  
 FT DOMAIN 650 661 EF\_HAND 3.  
 FT DOMAIN 680 691 ANCESTRAL CALCIUM SITE 4.  
 FT ACT\_SITE 105 105 ANCESTRAL CALCIUM SITE 5.  
 FT ACT\_SITE 262 262  
 FT ACT\_SITE 286 286  
 FT MUTAGEN 105 105 C->S: LOSS OF ACTIVITY.  
 FT MUTAGEN 226 226 K->S: DECREASE OF 12% OF THE ACTIVITY.  
 FT MUTAGEN 230 230 K->S: NO EFFECT.  
 FT MUTAGEN 230 230 K->E: DECREASE OF 84% OF THE ACTIVITY.  
 FT MUTAGEN 234 234 K->E: DECREASE OF 85% OF THE ACTIVITY.  
 FT MUTAGEN 234 234 K->S: DECREASE OF 20% OF THE ACTIVITY.  
 FT MUTAGEN 262 262 H->A: LOSS OF ACTIVITY.  
 FT MUTAGEN 286 286 N->A: LOSS OF ACTIVITY.  
 FT MUTAGEN 288 288 W->Y: DECREASE OF 95% OF THE ACTIVITY.  
 FT MUTAGEN 504 504 E->S: DECREASE OF 10% OF THE ACTIVITY.  
 SQ SEQUENCE 700 AA; 79919 MW; 296B0DC3BEEF5B90 CRC64;

Query Match 43.3% Score 45.5; DB 1; Length 700;  
 Best Local Similarity 47.4%; Pred. No. 36;  
 Matches 9; Conservative 6; Mismatches 3; Indels 1; Gaps 1;  
 QY 1 KWKLLKKPPLKLLKK 19  
 DB 213 EWE-LKKPPLKLLKK 230

CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>  
 CC or send an email to [license@sdb.ch](mailto:license@sdb.ch)).  
 CC -----  
 DR EMBL; X17205; CAA35077.1; -  
 DR FIR; S07883; S07883.  
 KW Hypothetical protein; Plasmid.  
 SQ SEQUENCE 151 AA; 16744 MW; B20A0C57913DCAB CRC64;

Query Match 42.9% Score 45; DB 1; Length 151;  
 Best Local Similarity 66.7%; Pred. No. 9.1;  
 Matches 8; Conservative 3; Mismatches 1; Indels 0; Gaps 0;  
 QY 8 KPPPLKLLKK 19  
 DB 108 EPPPLKLLKK 119

Search completed: June 6, 2003, 13:24:00  
 Job time : 9 secs

RESULT 15  
 YP16\_METTM  
 ID YP16\_METTM STANDARD; PRT; 151 AA.  
 AC P14935;  
 DT 01-APR-1990 (Rel. 14, Created)  
 DT 01-APR-1990 (Rel. 14, Last sequence update)  
 DT 15-JUL-1998 (Rel. 36, Last annotation update)  
 DE Hypothetical 16.7 kDa protein.  
 OS Methanobacterium thermoautotrophicum (strain Marburg / DSM 2133).  
 OG Plasmid pME2001.  
 CC Archaea; Euryarchaeota; Methanobacteria; Methanobacteriales;  
 CC Methanobacteriaceae; Methanothermobacter.  
 CX NCBI\_TaxID=79929;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=90221821; PubMed=2326168;  
 RA Bokranz M., Klein A., Meile L.;  
 RT "Complete nucleotide sequence of plasmid pME2001 of Methanobacterium  
 RT thermoautotrophicum (Marburg)."  
 RL Nucleic Acids Res. 18:363-363(1990).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its

